Departamento de Tecnología y Química Farmacéuticas
Facultad de Farmacia y Nutrición
UNIVERSIDAD DE NAVARRA

DEVELOPMENT AND IMPLEMENTATION OF NOVEL METHODOLOGIES TO IMPROVE PHARMACOMETRICS AND SYSTEMS PHARMACOLOGY ANALYSIS

Itziar Irurzun Arana
Pamplona, 2019

Departamento de Tecnología y Química Farmacéuticas

Facultad de Farmacia y Nutrición

UNIVERSIDAD DE NAVARRA

TESIS DOCTORAL

DEVELOPMENT AND IMPLEMENTATION OF NOVEL METHODOLOGIES TO IMPROVE PHARMACOMETRICS AND SYSTEMS PHARMACOLOGY ANALYSIS

Trabajo presentado por Itziar Irurzun Arana para la obtención del Grado de Doctor

Fdo. Itziar Irurzun Arana

Pamplona, 2019

UNIVERSIDAD DE NAVARRA

Facultad de Farmacia y Nutrición
Departamento de Tecnología y Química Farmacéuticas

D. JOSÉ IGNACIO FERNÁNDEZ DE TROCÓNIZ FERNÁNDEZ, Doctor en Farmacia y Catedrático del Departamento de Tecnología y Química Farmacéuticas.

Certifica:

> Que el presente trabajo, titulado "Development and implementation of novel methodologies to improve Pharmacometrics and Systems Pharmacology analysis" presentado por DÑA. ITZIAR IRURZUN ARANA para optar al grado de Doctor en Farmacia, ha sido realizado bajo su dirección en los Departamentos de Tecnología y Química Farmacéuticas. Considerando finalizado el trabajo autorizan su presentación a fin de que pueda ser juzgado y calificado por el Tribunal correspondiente.

Y para que así conste, firma la presente:

Fdo.: Dr. José Ignacio F. Trocóniz
Pamplona, 2019

A mi aita

“Nunca te transformes en una brisa suave
cuando naciste para ser tormenta.”

- Ron Israel

Agradecimientos

Quisiera comenzar expresando mi gratitud a la Universidad de Navarra y al Departamento de Tecnología y Química Farmacéuticas por haberme posibilitado la realización de esta tesis doctoral.

En segundo lugar deseo expresar todo mi agradecimiento al director de esta tesis, el Profesor Iñaki Trocóniz, por aceptar mi petición de formar parte de su equipo durante estos cuatro años. Muchas gracias por tu apoyo y confianza en mi trabajo, por dar valor a mi opinión y por haberme dado la oportunidad de asistir a congresos internacionales y hacer una estancia predoctoral en Boston. Considero que has sido un gran mentor y espero sinceramente no perder tu contacto con el paso de los años. Sobre todo porque queda un chuletón pendiente.

I would also like to thank Prof. Franziska Michor for giving me the opportunity of joining her group for 6 months in the Department of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute (Boston, USA). My sincere thanks also goes to the rest of the members from the Michorlab, especially to Ollie, Shaon, Danny, Jamie and Kam. My internship in the department has been a period of intense learning, not only in the scientific arena, but also on a personal level. Thanks for being part of it.

Sin embargo, de este periodo en la Universidad de Navarra no solo me llevo una tesis doctoral, sino también unas grandes amigas, las pkpditas, María, Leire, Belén y Violeta, sin duda sin vosotras esto no hubiera sido lo mismo. Gracias por vuestra amistad y apoyo moral, por los viajes que hemos disfrutado juntas aprovechando conferencias, por Creta, Lisboa, Budapest, San Diego y los que nos queden. Por grandes aventuras, risas y meteduras de pata que siempre nos recordaremos mutuamente. No estar junto a vosotras es algo a lo que me va a costar acostumbrarme.

Gracias a las últimas incorporaciones del grupo, Diego y Nico. Sé que mantendréis el nivel de risas, cenas y karaokes que hemos construido durante estos años. Enseñad a los nuevos que para formar parte del grupo no basta con trabajar correctamente, sino que hay que implicarse emocionalmente, hay que organizar actividades y es obligatorio subir al café. Confío en vosotros.

Deseo también expresar mi más sincero agradecimiento a grandes postdocs y profesores con los que he tenido la suerte de coincidir, Edu, Zinnia, Nuria, María Jesús, Aitor, Victor y JD, gracias a las consultas, debates y aclaraciones que he podido compartir con vosotros, siempre habéis sido de gran ayuda para reforzar mis conocimientos y he podido contar con vosotros en todo momento.

Tampoco me olvido de las personas que han venido de estancia por unos meses al departamento, especialmente gracias a Nacho, Aziz, Aurelia y Fernando por integrarse tan bien en el grupo. Sois grandes personas a las que me alegra mucho haber conocido y que hayáis formado parte de esta experiencia. Espero que podamos volver a coincidir en el futuro.

Quisiera mencionar también a mis ex compañeras de universidad Maitena y Jaione, gracias por esas cenas de viernes en las que me habéis dado la oportunidad de desahogarme y de la capacidad de comprensión que tenemos entre nosotras al ser las únicas ingenieras biomédicas de nuestros respectivos grupos. ¡Ánimo que a vosotras tampoco os queda tanto!

Gracias también a mi kuadrilla (sí con k), por las cenas en la soci, casas rurales, los audios vergonzosos que os encargáis de recopilar, por nuestra obsesión con Donosti y poder estar allí el día de la tamborrada, por tener la capacidad de reírnos de los defectos de cada uno de nosotros sin que nadie se moleste, porque somos muy brutos, pero ante todo, somos leales.

Gracias a David, porque siempre me recuerdas que al final todo esto es trabajo y que lo más importante en la vida está fuera de aquí. Perdóname si a veces se me olvida. Gracias por ayudarme a restar importancia a cosas que realmente no la tienen tanto y por ayudarme a ser mejor persona siempre.

Por supuesto, estaré siempre eternamente agradecida a mi familia, especialmente a mi Ama, mi tío Eusebio y mi tía Arantxa, por sufrir mis decepciones y por disfrutar mis alegrías. Mención especial se merece mi madre, por ser el ejemplo de mujer luchadora y trabajadora en el que me baso y escucharme siempre que lo necesito. Por tu ilusión cuando consigo lo que me propongo y enmarcar todos mis títulos y premios con orgullo.

Y por último mencionar a una persona muy especial que aunque ya no esté conmigo, sé que de estarlo, estaría muy orgulloso de mí. Mi Aita. Porque tus consejos siempre eran los mejores, por enseñarme a disfrutar de la vida y animarme a perseguir mis sueños. Me enseñaste que el mundo está dirigido por los que se dan a conocer, no por los que esperan a que se pregunte por ellos. Sé que me hubieras seguido al fin del mundo y espero que allí, algún día, nos encontremos de nuevo. Eres el adiós que jamás sabré decir y te dedico, especialmente a ti, todo el fruto del trabajo de estos cuatro años de tesis doctoral.

A todos, muchas gracias. Compartir este tiempo con vosotros es algo que nunca olvidaré.

Table of contents

Abbreviations

AGN	Endogenous agonist
AJCC	American Joint Committee on Cancer
AUC	Area Under the Curve
BFs	Boolean functions
BOV	Between-occasion variability
BSV	Between-subject variability
CL	Total drug clearance
CL_D	Distribution clearances between the central and peripheral compartment
CT	Castration limit
CV	coefficient of variation
D_{inf}	Duration of the zero-order process
D_R	Down-regulation process
ECOG	Eastern Cooperative Oncology Group
Emax	Maximum effect
FDA	Food and Drug Administration
FDB	Feedback
GAM	Generalized Additive Models
IFN-$\alpha 2b$	Interferon $\alpha 2b$
K_A	First-order rate constant of the depot compartment
k_{D_T}	First-order rate constants of testosterone degradation

K_D	Receptor equilibrium dissociation constant
k_{S_R}	Zero-order rate constants of receptor synthesis
LDH	Lactate Dehydrogenase
M&S	Modeling and simulation
mAbs	monoclonal antibodies
MIA	Melanoma-inhibiting activity
MID3	Model Informed Drug Discovery and Development
NLME	Nonlinear mixed-effects
NPC	Numerical predictive checks
OC	Optimal Control
ODE	Ordinary Differential Equation
PD	Pharmacodynamics
PI	Perturbation Index
PK	Pharmacokinetics
popPKPD	Population Pharmacokinetics and Pharmacodynamics
PSP	Pharmacometrics and Systems Pharmacology
QSP	Quantitative systems pharmacology
RV	Residual variability
SBML qual	Systems Biology Markup Language Qualitative Models
SDE	Stochastic differential equation
SP	Systems Pharmacology
SPIDDOR	Systems Pharmacology for effIcient Drug Development On R
SR	Sustained-release
t_{cast}	Time to achieve testosterone castration limit values
t_{effect}	Castration time after drug injection

TRP	Triptorelin
TST	Testosterone
TST_0	Baseline testosterone level
TST_{max}	Maximum testosterone level
V	Apparent volume of distribution
VPC	Visual predictive checks

Preface

During the past decades Pharmacometrics and Systems Pharmacology (PSP) modelling has emerged as a promising discipline within drug development context. Model-based approaches in drug development involve the integration of pharmacokinetics (PK), pharmacodynamics (PD), disease progression and other relevant information to describe complex biological systems and the action of drugs by computational models. The use of such models can have a major impact during all phases of drug discovery and development and may ultimately result in significant cost reductions for the pharmaceutical industry.

Modeling and simulation (M&S) in PSP integrates diverse scientific domains including pharmacology, mathematics, computer science, biostatistics, systems biology, and recently even artificial intelligence is being applied in this field. The diversity of this discipline sometimes results in the challenge that people of different backgrounds do not share the same knowledge about the different aspects governing M&S arena.

The present thesis explores the possibility to improve standard PSP modelling by integrating different methodologies and tools that can aid to build a bridge between the different disciplines in order to develop more mechanistic pharmacological models.

This thesis is structured as follows:

The *Introduction* section gives a general overview about the different methodologies that will be explored in the different chapters of this thesis.

Chapter 1 proposes a qualitative modeling strategy which consist on a computational framework to perform simulations of Boolean networks in the R environment and analyze the result of the perturbations on these networks. This framework called SPIDDOR (from Systems Pharmacology for effIcient Drug Development On R) combines the advantages of the parameter-free nature of logical models while providing a good approximation of the qualitative behavior of pharmacological systems, making the use of Boolean networks in SP more accessible to scientist involved in drug development, especially at its early stages. Additionally, this tool has been used to qualitatively evaluate the results of Boolean network models describing pathogenic mechanisms in the autoimmune diseases systemic lupus erythematosus and inflammatory bowel disease. The publications corresponding to these works are added in the *Appendix* of the thesis.

Chapter 2 proposes an optimization technique known as Optimal Control and its application to a PKPD model for the testosterone effects of triptorelin, a synthetic gonadotropin-releasing hormone analog used to induce chemical castration in prostate cancer patients, with the goal of improving the release characteristics of the drug. As the proposed approach is not circumscribed to just this particular problem, the reader will find a comprehensive description of how the critical aspects of defining control variables and selecting the cost functions and constraints were handled.

Chapter 3 presents a computational framework based on a stochastic model known as *multitype branching process* used to explore the dynamic evolution of heterogeneous tumor cell populations. This framework, which also consist on an R package, is called ACESO (from A Cancer Evolution Simulation Optimizer) and incorporates pharmacokinetics and drug interaction effects into the stochastic model. The aim of this tool is to identify optimum dosing schedules that minimize the risk of developing resistance to anticancer therapies.

The need for integration of data from diverse sources including pharmacokinetics, pharmacodynamics, disease progression and toxicology has led also to an increased use of mechanism-based models during early phases of drug development but also in clinical cancer research. Therefore in *Chapter 4* a semi-mechanistic model describing the time course of several circulating biomarkers in advanced melanoma patients treated with adjuvant high-dose interferon $\alpha 2b$ is presented in order to evaluate the dynamics of the tumor markers as prognostic factors of the overall survival and progression-free survival of the patients. This treatment-biomarker-survival model is also coupled to another semi-mechanistic model describing the side effects of interferon therapy in the absolute neutrophil counts of the patients in order to simultaneously analyze the benefits and toxic effects of this treatment.

Finally, the *General Discussion* integrates and highlights the most relevant aspects of the four chapters, to end with the *Conclusions* where a summary of the more relevant findings of the thesis are described.

Introduction

Introduction

Current drug discovery and development, regardless of the therapeutic area, is a long and complex process associated to a high attrition rate which results in an unacceptable high cost-benefit ratio going beyond the drug industry context. This fact motivated the emergence of a strategic initiative by the Food and Drug Administration (FDA) in 2004, the critical path initiative, with the goal of optimizing the development programs and accelerating the drug approval process [79].

Considering the need to anticipate/extrapolate drug safety and efficacy in new scenarios based on information gathered at previous stages (i.e., from preclinical to first dose in humans, or from healthy volunteers to patients) and its uncertainty associated, is not surprising that one of the six broad topic areas of the mentioned critical path initiative was devoted to promote the application of mathematics, statistics and computational analysis within the drug development and regulatory review process. That represented the first step towards the current Model Informed Drug Discovery and Development (MID3) paradigm enclosed within the discipline of Pharmacometrics and Systems Pharmacology (PSP).

Mathematical modeling and simulation is frequently used in this area as it provides a low-cost method to identify and integrate major determinants of drug action and make predictions about disease progression and/or drug effects that could inform key decisions in the drug development process. PSP, from a high level perspective, can be viewed as a way to integrate and quantitatively link all type of information, methodologies and tools associated with the development of mechanistic models of biological and/or physiological processes and pharmacology that can help in our understanding and subsequent prediction of drug response in patients. Essentially, pharmacometrics is defined as "the science of developing and applying mathematical and statistical methods to characterize, understand, and predict a drug's pharmacokinetic (PK) and pharmacodynamic (PD) behaviour" [78]. Systems Pharmacology (SP) on the other hand, is an emerging multi-scale discipline that focuses on the combination of the interactions among multiple levels of biological organization (molecules, cells, tissues, organs, etc.) and basic principles of PKPD as a means to describe and predict therapeutic and adverse drug effects at the whole-organism level [66].

The value of MID3 approaches in enabling model-informed decision-making is evidenced by a large number of publications. Some recent examples can be found in [28, 56, 67, 8].

In these examples, the reliability and explanatory power of the chosen models depend predominantly on the level of detail included and the assumptions they are based on. In many cases a combination of strategies is the best way to proceed. Especially in pharmacometrics, not only quantifying the mean tendency of a population with a deterministic function is of particular interest, but also finding ways of modelling the unexplained variance of the data, which may be the result of differences in patient

characteristics and/or uncontrolled sources of noise caused by the inaccuracy of the analytical methods or errors in sampling times. This is the case of nonlinear mixed effects modeling approach (more commonly known as population pharmacokinetic/pharmacodynamic (popPK/PD) modeling), extensively used by the pharmaceutical industry and regulatory agencies to analyze experimental and clinical data and provide tools to design new scenarios and find optimal dosing schedules that extract the most from preclinical and clinical studies. Nevertheless, in clinical phases of drug development, these procedures are still mainly driven by classical maximum tolerated dose (MTD) schemas and non-compartmental analysis. Combination therapies present additional challenges where computational modeling can provide guidance. Because of the complexities of dose combinations and schedule options for therapy consisting of multiple agents, testing all combinations of schedules and doses is not possible. The use of mathematical models can help in the selection of optimal drug combinations, but even using these strategies, optimization of drug exposure is not a trivial task (especially in the case of non-linear and complex biological systems) and cannot be efficiently addressed through trial-error simulation exercises. In those cases, complementary optimization techniques that will be discussed later in this review need to be employed.

It is remarkable that the efforts performed in this field have resulted in the integration into the MID3 paradigm of very different modelling approaches that in the past were not connected at all to biopharmaceutics, pharmacokinetics or pharmacodynamics, as it is the case of optimal control [46] methods and the network models based on Boolean operators [32]. These are, however, areas and quantitative approaches that are seldom used by pharmacometricians and which deserve consideration and work to a mutual convergence. Figure 1 represents all the disciplines evaluated in this thesis and that could provide important insights into PSP. One of those cases is the stochastic modelling approach where, opposed to the most often used deterministic systems, the same set of parameter values and initial conditions can lead to different outcomes of the system. Stochasticity is very appealing as events occurring at random can have an important repercussion on disease progression and treatment effects. Currently, this mathematical approach is being used to quantify relevant pathological aspects such as the growing population of cancerous cells and bacterial or viral infections were mutations leading to drug resistant cells or individuals occur at random [16, 33].

MID3 has proven to be an extremely valuable tool to evaluate different treatment strategies more efficiently during the different preclinical and clinical drug development phases, but also to improve patient care as it narrows the set of possible schedules that could be tested in clinical settings that maximally extend patient survival. In this introductory review, we aim to give a general overview about each of the above mentioned methodologies, leaving the discussion about qualitative network models for Chapter 1.

The rest of the introduction is organized as follows: first a general description about deterministic and stochastic modeling strategies is provided highlighting the main differences in their model structure,

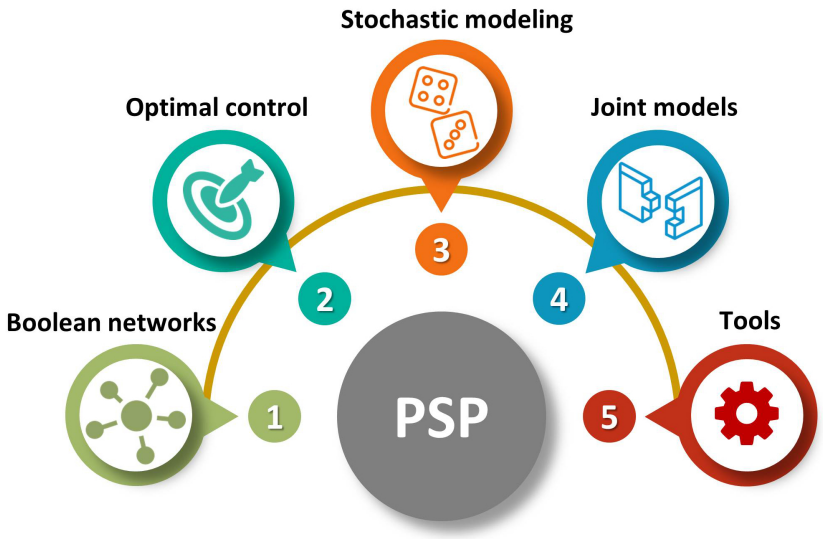


Fig. 1 Different type of modeling elements and techniques that can contribute to Pharmacometrics and Systems Pharmacology (PSP) area.

the simulation methods and model fitting techniques. Then the nonlinear mixed-effects modeling approach is presented as the most appropriate method to model the data coming from different individuals. Finally, the methodology of optimal control is introduced and examples of its application to deterministic and stochastic models are described. Each of these sections are complemented by the relevant literature in the field and examples of possible applications.

Deterministic vs stochastic models

Model structure

The structure of a model describes the relationship between the input variables and the output of a system. This relationship can be based on different mathematical formalisms depending on the degree of detail and the scope of the modeling effort, ranging from qualitative strategies such as Boolean models to quantitative approaches based on ordinary differential equations or stochastic processes. In this introduction we will focus on quantitative modeling strategies.

Quantitative models can be further divided into deterministic and stochastic systems. Deterministic models are often described by a system of algebraic or ordinary differential equations (ODEs) and describe the evolution of a system in continuous time. Here, the output of the model is fully determined by the parameter values and the initial conditions of the system and no randomness is assumed to be

present. Thus, the outcome of a deterministic model is unique and represents the average behaviour of the system. This average behavior is in most cases a suitable representation of real biological processes.

For instance, consider a cell population n where each cell gives birth to an offspring according to a constant birth rate β or dies according to a certain death rate δ. Assuming that the population grows exponentially then the ODE governing the exponential growth of the population can be defined as:

$$\frac{dn(t)}{dt} = (\beta - \delta) \cdot n \tag{1}$$

The analytical solution of this equation is: $n(t) = n_0 \cdot e^{(\beta-\delta)\cdot t}$.

This curve is fully determined by the birth and death rates and the initial condition of the population n0 : if the birth rate exceeds the death rate, $\beta > \delta$, the population size exponentially increases and when $\beta < \delta$, it exponentially decreases towards zero.

Tumor growth inhibition models are more elaborated examples of deterministic models used to analyze the effect of anti-cancer agents in oncology [34, 52, 5, 58]. In these models the change in tumor size is generally explained by the net tumor growth minus tumor shrinkage due to drug effects. Tumors are often assumed to growth exponentially or following the so-called logistic curves or Gompertz law [24]. In [51], Norton and Simon proposed a tumor growth following a Gompertz model where drug effects produce a decrement in the proliferation rate of the tumor. They found that the rate of cell killing by many drugs was proportional to tumor growth rates and they suggested that the best way of treating the patients was to increase the intensity of the treatment as the tumor became smaller [52, 50]. In a more recent study, García-Cremades and coauthors described pancreatic tumour progression by an exponential growth model and response to treatment by a linear decrease caused by the weekly area under the curve (AUC) of the drug plus a kinetic term reflecting reversible resistance to treatment [22] .

The limitation of deterministic models is that they do not account for uncertainty in model dynamics. This limitation often poses a practical problem because biological processes are always subject to stochastic effects and are intrinsically heterogeneous.

In stochastic processes it is assumed that each biological/pharmacological process is a random event that can take place with a certain probability related to the properties of the components within the biological system. Thus, the evolution of the system over time is dependent on a series of consecutive probabilistic events.

The stochastic approach needs a larger mathematical foundation than the deterministic approach, particularly in probability theory. The master equation is the fundamental mathematical description of this approach and the starting point to analyze a particular model. It is obtained as a probability

distribution for all the events that can occur during the time interval $(t, t+\Delta t)$, Δt being a very short time interval.

To continue with the example of the cell population birth-death process described above, we consider the stochastic version of the previous model where we assume that the time interval Δt is short enough so that only one of the following cases can exclusively occur: 1) a new cell is born with probability $\beta\Delta t$, 2) the cell dies with probability $\delta\Delta t$, and 3) the cell neither gives birth nor dies with probability $1-\beta\Delta t-\delta\Delta t$. Now we need to calculate the probability distribution $P_n(t)$, i.e. the probability that the population size is n at time t, which is the sum of the probabilities of all the mutually exclusive events explained above and is given as (for a more detailed explanation of this point see the lecture from [71]):

$$P_n(t+\Delta t) = P_n(t)(1-\beta n\Delta t-\delta n\Delta t)+P_{n-1}(t)\beta(n-1)\Delta t+P_{n+1}(t)\delta(n+1)\Delta t \tag{2}$$

By letting $\Delta t \to 0$, we obtain the master equation of the birth-death process:

$$\frac{dP_n(t)}{dt} = \beta(n-1)P_{n-1}(t)+\delta(n+1)P_{n+1}(t)-(\beta+\delta)nP_n(t) \qquad for \quad n \geq 1 \tag{3}$$

This equation allows us to compute the average and variance population size values by solving the first and second moments of the master equation:

$$Average: \quad n = n_0 \cdot e^{(\beta-\delta)t} \tag{4}$$

$$Variance: \quad Var[n] = n_0 \cdot \frac{\beta+\delta}{\beta-\delta} \cdot e^{(\beta-\delta)t} \cdot (e^{(\beta-\delta)t}-1) \tag{5}$$

We note that the analytic equation obtained for the average of the stochastic process is the same as the eq. 1 of the deterministic model. Many deterministic models represent the average behavior of all possible stochastic evolutions, that is, the differential equations of deterministic modeling are nearly equivalent to the differential equation that describes the time evolution of the mean of the stochastic process.

Unfortunately, the analytical solution of a master equation is often intractable, especially if a large number of components are involved. The Gillespie algorithm provides exact simulations of trajectories of the master equation [23], but the computational cost can be high for large biological systems because simulations must be repeated many times to reveal the complete range of behaviors of the system. That is one of the reasons why most biological processes in biomedical research are defined using deterministic models where the simulation of the system can be easily performed using the high number of numerical ODE solvers available. Some stochastic simulation software tools exist for easier computation of these processes like StochSS [17], GillesPy [1] or COPASI [30] but a great amount of works about stochastic process simulation in biomedicine found in the literature are still

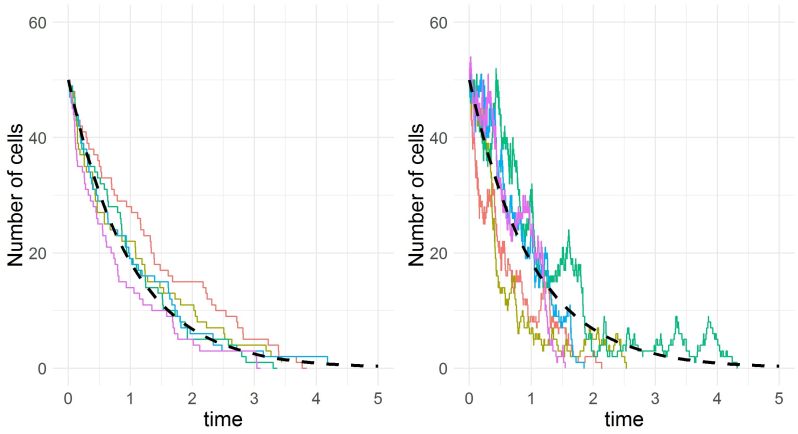


Fig. 2 Five realizations of a stochastic linear birth-death process together with the continuous deterministic solution for two different birth and death rate combinations, each with $\beta - \delta = -1$ (arbitrary units)and an initial number of cells of 50 (left: $\beta = 0$ and $\delta = 1$; right: $\beta = 4$ and $\delta = 5$).

accomplished using self-coded algorithms in programming languages like C++ or Mathematica due to the fast computing time needed.

In Figure 2, five stochastic realizations of the birth-death process together with the continuous deterministic solution for two different birth and death rate value combinations are shown. Under the deterministic model, the number of cells never reaches zero, simply tends to zero as time goes to infinity. In the stochastic simulations however, the cells go extinct and there is considerable randomness associated with the time when this occurs. The possibility that the population of cells can die out is an important feature of stochastic models opposed to deterministic ones. Another point to highlight is that the stochastic process depends explicitly on both the birth rate and the death rate, and not just on the net growth rate $(\beta - \delta)$ like in the case of the deterministic model. As shown in Figure 2, the shape of the curve is governed by result of $(\beta - \delta)$, but the variance of the process is also controlled by the quantity $(\beta + \delta)$ as can also be seen from equation 4, thus the right panel shows a greater degree of noise than the left panel.

Deterministic models provide good approximations to the growth of large populations but for small populations it is important to take into account their inherent randomness because these populations can go extinct due to random fluctuations even in the case where the birth rate of the population is greater than the death rate, something impossible when considering deterministic models. That is why many authors focused on the calculation of the probability of extinction of a population [11, 21, 3]. For the single-type birth-death process previously described, the probability of extinction of the cell population is given by (see [2] for a more exhaustive mathematical explanation):

$$P(n(t) = 0) = \left(\frac{-\delta \cdot e^{(\delta - \beta)t} + \delta}{-\delta \cdot e^{(\delta - \beta)t} + \beta} \right)^{n_0} \tag{6}$$

From this equation we can see that when $t \to \text{inf}$ and $\beta < \delta$ the probability of extinction will be 1 but when $t \to \text{inf}$ and $\beta > \delta$, the probability of extinction becomes $P(n(t) = 0) \sim \left(\frac{\delta}{\beta}\right)^{n_0}$, which means that for a small population size n_0 we can obtain a probability of extinction of the process different from 0. This property introduces the possibility of a complete cure of the patient if the tumor cells become extinct caused by the effect of the anti-cancer treatment.

As mentioned in the introduction, another important characteristic of stochastic models is that the same initial conditions and parameter values can lead to different behaviour of the system. For example, in [3] they propose a stochastic mathematical model for the analysis of adoptive cell transfer immunotherapy against melanoma skin cancer where the treatment consist on the injection of T-cells that recognize a melanocyte-specific antigen and are able to kill differentiated melanoma cell types. In this model, the authors set the initial conditions such that the number of differentiated melanoma cells is large, the number of T-cells injected is small and the initial number of dedifferentiated melanoma cells (not sensitive to T-cell therapy) is small too. T-cells control the growth of the differentiated melanoma cells but, due to the random fluctuations of the system, the small number of T-cells injected can die out, which will lead to the growth of the differentiated melanoma cells. This behavior would not be seen in the deterministic version of the same model. In addition, to introduce the possibility that both differentiated and dedifferentiated melanoma cells become extinct (the patient is cured), the authors suggested that two types of T-cells should be injected, one specific to the differentiated population and the other one specific to the dedifferentiated cells.

Stochastic models have been widely used to model clonal evolution in growing tumors and the evolution of resistance to anti-cancer drugs, which occurs when an initially sensitive tumor no longer responds to treatment due to randomly generated (epi)genetic alterations in cancer cells [54, 45, 37, 33, 25, 9]. Here, discrete stochastic models describe cell growth and mutation acquisition by defining probabilistic reaction rates of the respective events in the form of a Markov process (the probability of observing n cells at time $t + \Delta t$ only depends on the state of the population at time t and not on its earlier history). Some of the most popular models in this context include branching processes [27, 35, 18], and Moran processes [47].

Branching process

Branching processes are a class of stochastic models that describe the growth and composition of populations by stochastically reproducing individuals [35]. Multi-type branching processes (branching processes where offspring can be of different type than the parent) are convenient for modelling clonal evolution of cancer cells because new (epi)genetic alterations emerge as random events during cell division and give rise to tumor subclones with different fitness than their ancestors [54]. Here, each cell is fully described by cell-intrinsic birth, mutation and death rate. A schematic representation of this process is shown in Figure 3A.

Apart from simulating the growth of heterogeneous cancer cell populations, the distinctive characteristic about these models compared to the deterministic approach is that they allow for the calculation of the probability of developing resistance at any time interval once the treatment has begun [33]. This probability of resistance is 1 minus the result of the sum of the probability of no sensitive cell divisions giving rise to a resistant cell in the studied time interval plus the probability that the resistant cells become extinct.

Several authors have attempted to quantitatively model the response of cancer cells to chemotherapy or targeted therapies using this type of models assuming that resistance can be produced by random mutations or be present at the start of the therapy [37, 21, 10, 20]. In [10] the authors determined differential growth kinetics for drug-sensitive and drug-resistant EGFR-mutant non-small cell lung cancer cells during therapy with two tyrosine kinase inhibitors in order to predict optimum dosing strategies. In a subsequent study they incorporated PK processes into the stochastic model demonstrating the power of mathematical modeling in predicting improved treatment schedules with existing drugs [20]. A deterministic version based on these processes can also be found in [72]. Multitype branching processes are also powerful frameworks to analyze the data from barcoding experiments [57].

Moran process

The Moran process, named after the Australian statistician Pat Moran, is a widely-used stochastic model in population genetics [47]. In this model the total population size is fixed and stable coexistence of different cell types is impossible. At each time step, a cell is chosen to divide at random, but proportional to fitness. The chosen cell produces a daughter cell that replaces another randomly chosen cell that dies. Thus, the total number of cells remains strictly constant as can be seen in Fig 3B. This model is in general useful to describe cancer initiation since the total number of pre-cancer cells tends to be stable. In this early stage of cancer, Moran processes also allow slow accumulation of deleterious mutations while rapid accumulation of advantageous mutations [53].

Model fitting

When the structure of a mathematical model is defined and experimental data is available, the next step is to determine whether the model is capable of describing the experimental data, and quantify the parameter values that give a good fit. In clinical or preclinical trials for example, a typical longitudinal dataset consists on a number of observations (e.g. drug concentrations, biomarker level, tumor volume values...) collected among the different individuals participating in the trial.

Values of model parameters are obtained using estimation techniques aimed at searching the model that minimizes the difference between the observations in the datasets and the model prediction.

Extensive research has been done on numerical ODE solvers and advanced algorithms which allows for a straightforward estimation of parameters from deterministic models thanks to the development of easy-to-use tools. Stochastic models introduce additional challenges due to their intrinsic stochasticity and the level of computer science expertise needed to code the simulation algorithms, and therefore there is a lack of tools for easily finding parameter estimates for such models. A few frameworks like StochSS or COPASI exist to help the users to automatically complete this task but, even so, most studies found in literature treat the system as deterministic instead of stochastic when estimating the parameters of the model to be able to use common procedures like least-squares or maximum likelihood estimation. In those cases, the estimates of the parameter obtained through the deterministic formulation may be plugged back into the stochastic model for the simulation of the probabilistic process. A formal discussion of the theory of such algorithms is beyond the scope of this introductory review.

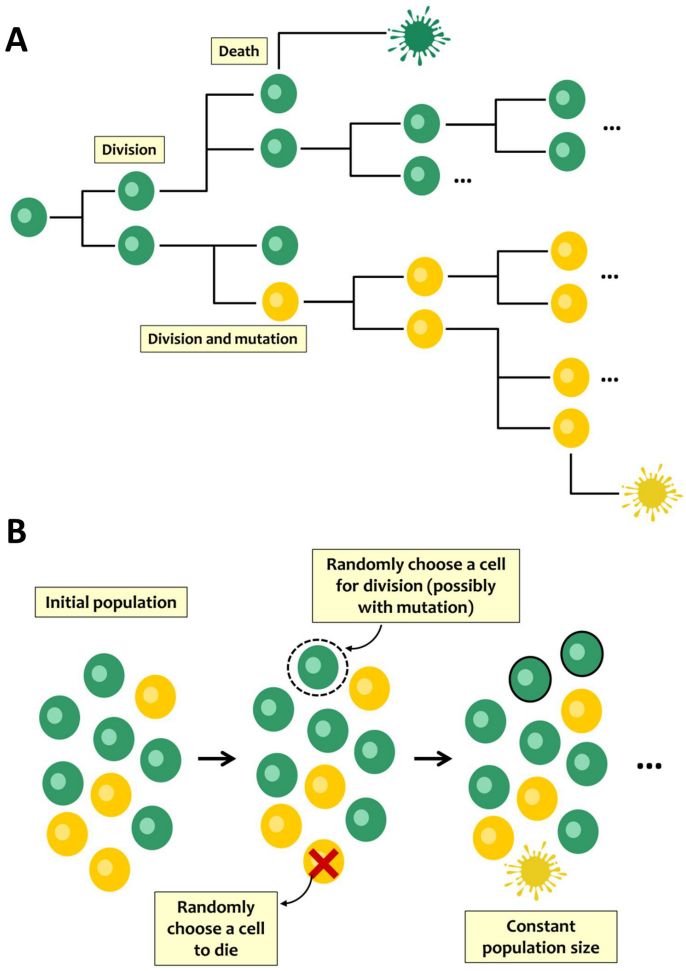


Fig. 3 Schematic representation of A) a two-type branching process, B) Moran process.

Even using deterministic approaches for parameter estimation, the model prediction won't fit perfectly to the data because there could be assay errors or simply because the model might not be completely accurate. This discrepancy is usually accounted for by a residual error term. However only defining this source of uncertainty might not be enough when modeling the data of different individuals. To account for a variation component in addition to the residual error, we introduce the nonlinear mixed-effects approach which combines the techniques of nonlinear regression and mixed-effects modeling.

Nonlinear mixed-effects models

Nonlinear mixed-effects (NLME) models were introduced into the pharmaceutical sciences to analyze data from several individuals simultaneously. This approach considers the population study sample rather than the individual and that is why this methodology is also known as the *population approach*. Here, the individuals are characterized by a common deterministic model known as the structural model and a statistical model which allows for model parameters to vary within the population. The term "mixed-effects" refers to the presence of both fixed effects that describe the typical parameter values of the population and random effects that handle the unexplained variability of the model. In general, there are two sources of random variability; variability assigned to model parameters and residual variability (RV) to account for the differences between the individual prediction of the model and the measured observation. Parameter variability can be further divided into between-subject variability (BSV) and between-occasion variability (BOV).

Mathematically, NLME models can be formally represented by:

$$y_{ij} = f(p_i, x_{ij}) + \varepsilon_{ij} \tag{7}$$

where y_{ij} is the observation obtained from the i^{th} individual at the j^{th} time, f represents a mathematical function that relates the vector of individual parameters p_i and the vector of independent variables x_{ij} (for example dose and time) to the observations and ε_{ij} is the model for the residual error which represents the discrepancy between model predictions and observations and is assumed to follow a normal distribution with mean 0 and variance–covariance matrix Σ. When f is nonlinear with respect to the parameters p, equation 7 describes the general form of nonlinear mixed effect models.

The vector of individual parameters p_i can be characterized by a function g of the fixed effects (p_{pop}), random effects (η_i) and covariates (Z_i) specific for each individual:

$$p_{ik} = p_{pop} \cdot e^{\eta_{ik}}$$
$$\eta_{ik} \in N(0, \Omega)$$

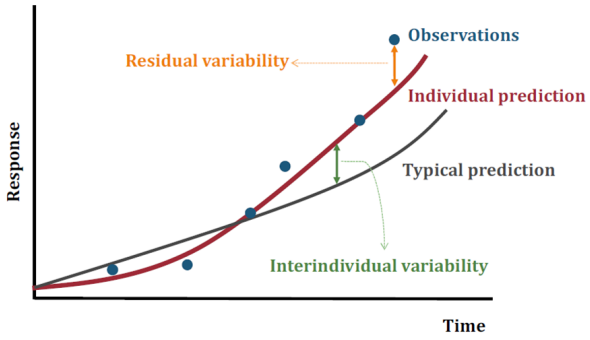


Fig. 4 Different type of variability in nonlinear mixed-effects models. Blue points represent observation values, gray line represents the typical prediction of the model and red line represents the individual prediction obtained for the observed data

where p_{ik} and p_{pop} represent the i^{th} individual and typical population values of the k parameter, respectively and η_{ik} corresponds to the deviation of p_{ik} with respect the typical value p_{pop}. Incorporation of random effects in NLME models is important because it permits prediction of the expected outcome and quantification of the magnitude of the variability and residual error. Figure 4 displays graphically the different type of variabilities along with the observations and individual and typical predictions of the model.

Covariate models describe the influence of intrinsic factors such as demographics (age, gender,...) or extrinsic factors such as smoking status, drug formulation, or concomitant medication on the individual time course of the response. Covariate inclusion may increase the predictive performance and mechanistic interpretability of mathematical models by, for example, identifying patient subgroups that may have suboptimal efficacy of the drug or be at increased risk due to the adverse events of the treatment.

The structural model in NLME approach is a deterministic model defined by an algebraic or ordinary differential equation. Even if some randomness in model parameters and errors is involved, this methodology does not allow for uncertainty in the dynamics of the underlying model. Therefore, this stochasticity must not be confused with the stochastic models explained earlier in this review. The stochastic component of NLME models is needed to describe the randomness involved in the measurements of the population but, given an initial condition and the parameter values for every individual, the output of the model will always be the same.

One way to introduce stochasticity in the dynamics of the model itself will be by using stochastic differential equations (SDE). NLME models based on SDEs extend the first-stage model of the hierarchical structure by decomposing the residual variability into measurement and system noise [73, 55]. Thus, here three levels of random-effects are included. In [73], the use of SDEs to characterize the

complex absorption properties of subcutaneous injections of the gonadotropin releasing hormone antagonist degarelix was evaluated. The authors suggested a state equation for the absorption half-life of the drug that fluctuates randomly. Tracking of the absorption half-life showed that after the third day of treatment, the absorption half-life started to increase and reaches a new constant value. This pattern was finally described using a usual NLME model with two absorption compartments, one with an initial fast absorption phase followed by a prolonged slow release phase in the second compartment.

Traditionally, the analysis of the variability in NLME model parameters was first studied using the standard two-stage approach (STS), where the estimates from each individual were obtained by the analysis of each individual separately and subsequently, the population parameters were derived by calculating descriptive statistics (e.g. mean and standard deviation of the model parameters). However, if the parameters for a specific individual are unreliably estimated or even unidentifiable due to missing data, it can lead to incorrect inference about the population. In contrast to STS analysis, the population approach offers the possibility of gaining integrated information from relatively sparse or unbalanced data. The most popular tools to estimate fixed and random effects parameters in compartmental NLME modeling are NONMEM [4] and MONOLIX [38] among others.

Pharmacometrics

NLME modeling is the gold standard methodology in pharmacometrics for the analysis of longitudinal pharmacokinetic and pharmacodynamic data collected in preclinical and clinical studies, especially in drug development. It provides a reasonable approximation of the dynamics of the drug in the body and of its effects. One of the first inclusions of mixed-effects modeling in biopharmaceutical research was pioneered in the 1970s by Sheiner and Beal [63, 62, 61]. Since then, pharmacometrics has been extensively used in a wide variety of therapeutic areas and in different stages of the drug development process and pharmacology patient management[59, 65, 76] .

Pharmacokinetics

Pharmacokinetics is currently defined as the study of the time course of drug absorption, distribution, metabolism, and excretion (ADME). It describes the time course of the concentration of a drug in a body fluid, generally plasma or blood, that results from the administration of a certain dosage regimen.

Two main approaches exist to characterize the PK of a compound: i) non-compartmental analysis (NCA), in which descriptive statistics such as maximum drug concentration (C_{max}), time to achieve C_{max} (T_{max}) or AUC are summarized directly from observed individual profiles; and ii) model-based compartmental analysis, in which the parameters governing the PK processes are inferred using nonlinear regression approaches. In NCA, total drug exposure is most often quantified by estimating the AUC of a concentration-time graph using the trapezoidal rule or other AUC estimating methods.

This method requires fewer assumptions than model-based approaches but it poses several limitations. First, summary statistics are generally only valid for a particular dose study, therefore, the prediction of untested dosing regimens is not generally possible if the PK data is nonlinear or the frequency of administration is to be adjusted. Second, this approach doesn't provide any insight into the underlying physiological processes that take place when a drug is administered to a patient. Compartmental models in contrast, use mathematical models (algebraic or ODE based) to estimate primary parameters (i.e., apparent volume of distribution, total clearance, absorption rate constant...) depending solely on the patient physiology, which enables to predict drug concentrations under different scenarios with the use of computer simulations.

Compartmental models view the body as relatively few, kinetically homogeneous compartments in which the drug is absorbed and distributed and from which elimination occurs. The simplest compartmental model is the one-compartment model. Here, the drug achieves instantaneous distribution throughout the body following administration and equilibrates instantaneously between tissues. Absorption and elimination processes are generally defined using first order rate constants. Thus the concentration-time profile shows a simple exponential decay (linear decay in log scale) following intravenous bolus drug administration (see Figure 5 for a schematic representation and model equations of a monocompartmental system following intravenous and extravascular administration).
Some drugs do not distribute instantaneously to all parts of the body though, even after intravenous bolus administration. In these cases, more compartments need to be included. The two-compartment model resolves the body into a central compartment and a peripheral compartment, which often represent extravascular less-perfused organs with nonrestrictive membranes (Figure 5 third panel). It is obvious that this approach represents a great simplification of the complex processes involved in our physiology, but for many compounds it has been proved to be sufficient to characterize drug exposure and establish appropriate therapeutic dosing regimens [74, 41, 42]. These models can be further complicated to allow for more than two compartments (Figure 5 last panel), nonlinear elimination or more complicated absorption profiles.

Pharmacodynamics

PD relates drug exposure to drug effects and, in conjunction with PK, allows characterizing the time course of in vivo drug action [29]. Hence, its objective is to relate the drug concentration at the site of action with the therapeutic effect of the drug.

Gerhard Levy was the first to develop a simple linear mathematical equation in the mid-1960s connecting the PK and the rate of decline of in vivo drug effects [39, 40]. Shortly thereafter, more elaborated direct effect models were proposed such as the so-called Emax model or sigmoid Emax model [77]. In this type of models, the drug immediately distributes to the site of action and therefore the pharmacological effects are readily observed and are directly related to the drug concentration.

However, generally, the pharmacologic response takes time to develop and the observed effects are not directly related to plasma concentrations of the drug. Reasons for these delayed responses may be attributed to different biological mechanisms like a slow distribution to the biophase, indirect mechanism of action, slow receptor deactivation or signal transduction/maturation, among others. Levy and

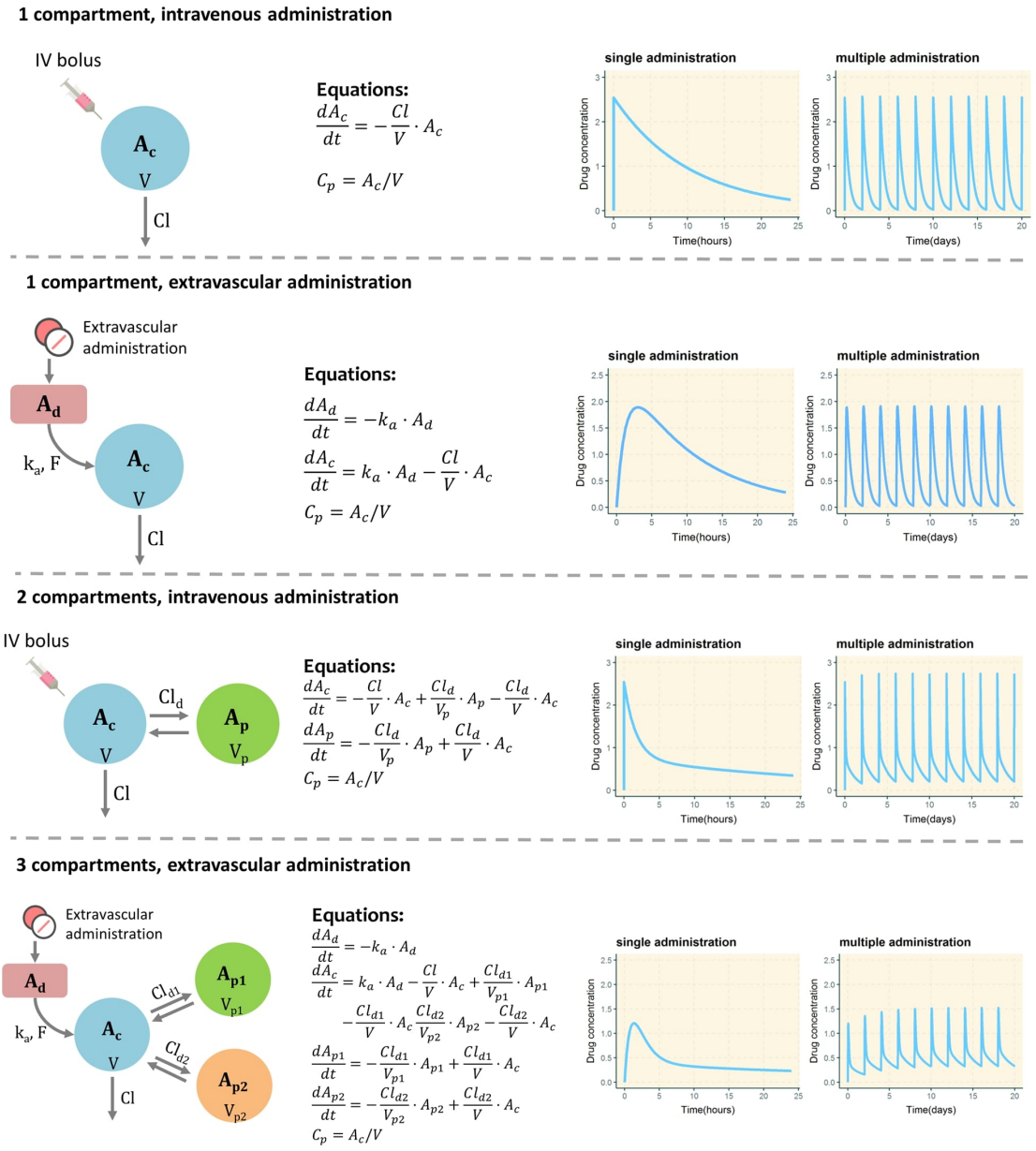


Fig. 5 Schematic representation and ordinary differential equations of different pharmacokinetic models. A_c, A_d and A_p represents the amount of drug in the central, depot and peripheral compartments respectively; CL is the apparent total clearance; V, V_p, are the apparent volumes of distribution of the central and peripheral compartments respectively; CL_d is the distribution clearance between the central and peripheral compartments; k_a is the first-order absorption rate constant; F is the bioavailability and C_p represents the drug concentration in plasma.

colleagues proposed one of the earliest examples of an indirect response model which characterized the anticoagulant effect of warfarin [48]. Years later, similar structural models are still widely in use. In oncology for instance, indirect response model are typically used to describe the levels of tumor biomarkers released to the circulating blood by tumor cells [6, 7]. Another commonly found model to explain this phenomenon involves using the effect-compartment approach which assumes that the rate of drug distribution to and from the hypothetical effect site determines the rate of onset of the pharmacological response [64]. In [75] the authors used this approach to characterize the time delay between plasma levodopa levels and the clinical response among patients with Parkinson disease.

Good reviews of these and different mechanism-based pharmacodynamic models can be found in [43, 19].

Optimal Control

Once mathematical models are built from experimental and/or literature data, they can predict the dynamics of the system under different conditions through computer simulations. However, simulation exercises are not always effective to obtain the desired objectives due to the complexity of the systems. Finding a solution to those problems is a challenging task which requires a complementary strategy obtained from engineering disciplines and control theory known as optimal control [68].

An optimal control (OC) problem is a dynamic optimization problem in which the state of a system is linked in time to the application of a control function, which drives the system towards a desirable outcome by minimizing a cost function subject to operating constraints. In other words, the control variable is able to manipulate the system to have an optimal performance.

OC has a long and successful history of applications in engineering and economics but also has become an important issue in biomedical and pharmacological research. The major objective of pharmacometrics for instance, is to aid in the identification of safe and effective dosing regimens that maintain drug concentrations within a therapeutic range while minimizing drug toxicity. In complex systems however, optimization of drug exposure is not a trivial task and cannot be efficiently addressed through simple parameter tuning simulation exercises, especially when drug combinations are taking into account. Thus, this type of methodologies is crucial to advance the mathematical problem of therapeutic optimization.

In this context, the question of how to define the objective functions and how to quantify our therapeutic goals becomes crucial. In clinical cancer research the control typically represents the drug dosage or, in simplified models, the effect of the drug on normal and cancer cells. Since chemotherapeutic agents affect normal cells as well as cancer cells, the objective becomes to minimize the number

of cancer cells over a fixed therapy interval while keeping the toxicity to the normal tissue at an acceptable level. One of the earliest studies where chemotherapy treatment planning was defined as an OC problem was in the work from [70], where the authors applied this technique to improve the treatment administration of the bone cancer IgG multiple myeloma. Since then, a significant amount of effort has been devoted to identify the most effective chemotherapeutic administration regimens using this type of methods [69, 44, 60].

In contrast to the above deterministic approaches, OC can also be applied to stochastic models if approximate mathematical solutions of the problem are obtained. As an example, Coldman and Murray used a multi-type branching process to describe the growth of drug sensitive and resistant cell subpopulations and tried to optimize drug regimen that achieves the best response [12]. In order to incorporate toxic effects into the model, the authors included a normal cell population following Gompertzian growth and defined an objective function that maximizes the probability of tumor extinction times the probability of no toxicity appearing during treatment. The authors concluded that early intensification of therapy is beneficial in situations in which resistance is likely.

Several studies also include the role of the immune system into the model and add the maximization of the number of immune system cells during or/and at the end of the treatment period to the objective function of the optimization problem [36, 14, 26, 49]. In the work from De Pillis and Radunskaya [14] they created a mathematical model for combined chemo- and immunotherapies with the aim of achieving maximum tumor reduction. As the dosage of the different drugs administered played the role of the controls, this is now a more complex mathematical problem because the authors had to deal with multiple controls. They showed that adding the minimization of the number of tumor cells at the end of the treatment was enough to decrease the tumor cell population at that specific time but this objective function induced large oscillations in the population during treatment period. Therefore, in a subsequent study, they added the minimization of the total tumor cell population summed over the course of treatment and the maximum tumor burden achieved during the course of treatment to the objective function in order to decrease this oscillatory pattern [15].

Nevertheless, not all the studies applying optimal control in biomedical/pharmacological research are related to optimizing therapeutic protocols. In [31] the release characteristics of a synthetic gonadotropin-releasing hormone analog used to induce chemical castration in prostate cancer patients were optimized using OC methods. The therapeutic goals formulated were the minimization of the initial flare up of testosterone levels and the time to reach testosterone values below castration limit and the maximization of the castration period of the patients. This work will be discussed in the second chapter of the thesis.

Concluding remarks

In this introduction, an overview of different modeling strategies have been described with the purpose of familiarizing the reader with the various concepts that will be covered in the following chapters of this dissertation.

Here we want to point out that, considering that each method has its own advantages and disadvantages, the possibility of combining different strategies may result in computational models with a greater potential than its predecessors. NLME approach for example, is the result of the combination of different classes of methods as it incorporates a structural or deterministic component plus a statistical model that describes the variability in the observed data.

Traditionally, deterministic and stochastic models have been thought to be rivals or at least opposite procedures. However, we argue that these approaches are complementary and that they can be combined to create more informative models. A discussion about the combination of different modeling strategies can be found in the *General discussion* section. One already mentioned example is the work from [10, 20] where the pharmacokinetics of erlotinib was linked to the growth kinetics of a drug-sensitive and drug-resistant non-small cell lung cancer cell line and a stochastic model based on branching processes was used to explore the dynamic evolution of these cancer cells under erlotinib treatment.

Quantitative systems pharmacology (QSP) represents another area where different modeling methods coming from Systems Biology and Pharmacometrics has been merged. Here drug effects can be simulated as perturbations of biological networks in order to understand the fundamental interactions that drive disease progression and treatment response. As an example, in the beginning of this introduction we referred to a recent work where a QSP model of the human coagulation network was developed to predict the effects of commonly used anticoagulants on clotting factor levels [28].

We strongly believe that a greater awareness of the different modelling methods and how to combine them will make future models more versatile and useful. For this reasons, in this work we propose the incorporation of additional modeling strategies and tools to the MID3 arsenal and show some examples of their application, with the ultimate goal of increasing drug development successes across all therapeutic indications.

References

[1] Abel, J. H., Drawert, B., Hellander, A., and Petzold, L. R. (2016). GillesPy: A python package for stochastic model building and simulation. *IEEE Life Sci Lett*, 2(3):35–38.

[2] Adke, S. R. and Moyal, J. E. (1963). A birth, death, and diffusion process. *J. Math. Anal. Appl.*, 7(2):209–224.

[3] Baar, M., Coquille, L., Mayer, H., Hölzel, M., Rogava, M., Tüting, T., and Bovier, A. (2016). A stochastic model for immunotherapy of cancer. *Sci. Rep.*, 6:24169.

[4] Beal, S. L., Sheiner, L. B., Boeckmann, A. J., and Bauer, R. J. (1992). NONMEM users guides. *NONMEM Project Group, University*.

[5] Bernard, A., Kimko, H., Mital, D., and Poggesi, I. (2012). Mathematical modeling of tumor growth and tumor growth inhibition in oncology drug development. *Expert Opin. Drug Metab. Toxicol.*, 8(9):1057–1069.

[6] Buil-Bruna, N., López-Picazo, J.-M., Moreno-Jiménez, M., Martín-Algarra, S., Ribba, B., and Trocóniz, I. F. (2014). A population pharmacodynamic model for lactate dehydrogenase and neuron specific enolase to predict tumor progression in small cell lung cancer patients. *AAPS J.*, 16(3):609–619.

[7] Buil-Bruna, N., Sahota, T., M. Lopez-Picazo, J., Moreno-Jimenez, M., Martin-Algarra, S., Ribba, B., and Troconiz, I. F. (2015). Early prediction of disease progression in small cell lung cancer: Toward Model-Based personalized medicine in oncology. *Cancer Res.*, 75(12):2416–2425.

[8] Campagne, O., Mager, D. E., Brazeau, D., Venuto, R. C., and Tornatore, K. M. (2018). The impact of tacrolimus exposure on extrarenal adverse effects in adult renal transplant recipients. *Br. J. Clin. Pharmacol.*

[9] Chakrabarti, S. and Michor, F. (2017). Pharmacokinetics and drug interactions determine optimum combination strategies in computational models of cancer evolution. *Cancer Res.*, 77(14):3908–3921.

[10] Chmielecki, J., Foo, J., Oxnard, G. R., Hutchinson, K., Ohashi, K., Somwar, R., Wang, L., Amato, K. R., Arcila, M., Sos, M. L., Socci, N. D., Viale, A., de Stanchina, E., Ginsberg, M. S., Thomas, R. K., Kris, M. G., Inoue, A., Ladanyi, M., Miller, V. A., Michor, F., and Pao, W. (2011). Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Sci. Transl. Med.*, 3(90):90ra59.

[11] Coldman, A. J. and Goldie, J. H. (1983). A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci.*, 65(2):291–307.

[12] Coldman, A. J. and Murray, J. M. (2000). Optimal control for a stochastic model of cancer chemotherapy. *Math. Biosci.*, 168(2):187–200.

[13] Costa, M. I., Boldrini, J. L., and Bassanezi, R. C. (1995). Drug kinetics and drug resistance in optimal chemotherapy. *Math. Biosci.*, 125(2):191–209.

[14] De Pillis, L. G. and Radunskaya, A. (2001). A mathematical tumor model with immune resistance and drug therapy: an optimal control approach. *Comput. Math. Methods Med.*

[15] De Pillis, L. G. and Radunskaya, A. (2003). The dynamics of an optimally controlled tumor model: A case study. *Math. Comput. Model.*, 37(11):1221–1244.

[16] Dewanji, A., Luebeck, E. G., and Moolgavkar, S. H. (2005). A generalized Luria–Delbrück model. *Math. Biosci.*, 197(2):140–152.

[17] Drawert, B., Hellander, A., Bales, B., Banerjee, D., Bellesia, G., Daigle, Jr, B. J., Douglas, G., Gu, M., Gupta, A., Hellander, S., Horuk, C., Nath, D., Takkar, A., Wu, S., Lötstedt, P., Krintz, C., and Petzold, L. R. (2016). Stochastic simulation service: Bridging the gap between the computational expert and the biologist. *PLoS Comput. Biol.*, 12(12):e1005220.

[18] Durrett, R. (2015). Branching process models of cancer. In Durrett, R., editor, *Branching Process Models of Cancer*, pages 1–63. Springer International Publishing, Cham.

[19] Felmlee, M. A., Morris, M. E., and Mager, D. E. (2012). Mechanism-Based pharmacodynamic modeling. In *Methods in Molecular Biology*, pages 583–600.

[20] Foo, J., Chmielecki, J., Pao, W., and Michor, F. (2012). Effects of pharmacokinetic processes and varied dosing schedules on the dynamics of acquired resistance to erlotinib in EGFR-mutant lung cancer. *J. Thorac. Oncol.*, 7(10):1583–1593.

[21] Foo, J. and Michor, F. (2010). Evolution of resistance to anti-cancer therapy during general dosing schedules. *J. Theor. Biol.*, 263(2):179–188.

[22] Garcia-Cremades, M., Pitou, C., Iversen, P. W., and Troconiz, I. F. (2018). Predicting tumour growth and its impact on survival in gemcitabine-treated patients with advanced pancreatic cancer. *Eur. J. Pharm. Sci.*, 115:296–303.

[23] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361.

[24] Gompertz, B. (1825). XXIV. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. FRS &c. *Philosophical transactions of the Royal Society of London*, 115:513–583.

[25] Greaves, M. and Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, 481(7381):306–313.

[26] Gu, W. and Moore, H. (2006). Optimal therapy regimens for treatment-resistant mutations of HIV.

[27] Haccou, P., Haccou, P., Jagers, P., Vatutin, V. A., and . Vatutin, V. A. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.

[28] Hartmann, S., Biliouris, K., Lesko, L. J., Nowak-Göttl, U., and Trame, M. N. (2016). Quantitative systems pharmacology model to predict the effects of commonly used anticoagulants on the human coagulation network. *CPT Pharmacometrics Syst Pharmacol*, 5(10):554–564.

[29] Holford, N. H. G. and Sheiner, L. B. (1981). Understanding the Dose-Effect relationship. *Clin. Pharmacokinet.*, 6(6):429–453.

[30] Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006). COPASI—a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074.

[31] Irurzun-Arana, I., Janda, A., Ardanza-Trevijano, S., and Trocóniz, I. F. (2018). Optimal dynamic control approach in a multi-objective therapeutic scenario: Application to drug delivery in the treatment of prostate cancer. *PLoS Comput. Biol.*, 14(4):e1006087.

[32] Irurzun-Arana, I., Pastor, J. M., Trocóniz, I. F., and Gómez-Mantilla, J. D. (2017). Advanced boolean modeling of biological networks applied to systems pharmacology. *Bioinformatics*, 33(7):1040–1048.

[33] Iwasa, Y., Nowak, M. A., and Michor, F. (2006). Evolution of resistance during clonal expansion. *Genetics*, 172(4):2557–2566.

[34] Jusko, W. J. (1971). Pharmacodynamics of chemotherapeutic effects: dose-time-response relationships for phase-nonspecific agents. *J. Pharm. Sci.*, 60(6):892–895.

[35] Kimmel, M. and Axelrod, D. E. (2015). *Branching Processes in Biology*, volume 19 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY.

[36] Kirschner, D., Lenhart, S., and Serbin, S. (1997). Optimal control of the chemotherapy of HIV. *J. Math. Biol.*, 35(7):775–792.

[37] Komarova, N. (2006). Stochastic modeling of drug resistance in cancer. *J. Theor. Biol.*, 239(3):351–366.

[38] Lavielle, M. (2005). MONOLIX (MOdèles NOn LInéaires à effets mixtes). *MONOLIX group, Orsay, France*.

[39] Levy, G. (1964). Relationship between elimination rate of drugs and rate of decline of their pharmacologic effects. *J. Pharm. Sci.*, 53:342–343.

[40] Levy, G. (1966). Kinetics of pharmacologic effects. *Clin. Pharmacol. Ther.*, 7(3):362–372.

[41] Lu, J., Eppler, S., Wolf, J., Hamilton, M., Rakhit, A., Bruno, R., and Lum, B. (2006). Clinical pharmacokinetics of erlotinib in patients with solid tumors and exposure-safety relationship in patients with non–small cell lung cancer. *Clinical Pharmacology & Therapeutics*, 80(2):136–145.

[42] Lu, J.-F., Bruno, R., Eppler, S., Novotny, W., Lum, B., and Gaudreault, J. (2008). Clinical pharmacokinetics of bevacizumab in patients with solid tumors. *Cancer Chemother. Pharmacol.*, 62(5):779–786.

[43] Mager, D. E., Wyska, E., and Jusko, W. J. (2003). Diversity of mechanism-based pharmacodynamic models. *Drug Metab. Dispos.*, 31(5):510–518.

[44] Martin, R. and Teo, K. L. (1994). *Optimal Control of Drug Administration in Cancer Chemotherapy*. World Scientific.

[45] Michor, F., Iwasa, Y., and Nowak, M. A. (2004). Dynamics of cancer progression. *Nat. Rev. Cancer*, 4(3):197–205.

[46] Moore, H. (2018). How to mathematically optimize drug regimens using optimal control. *J. Pharmacokinet. Pharmacodyn.*, 45(1):127–137.

[47] Moran, P. A. P. (1958). Random processes in genetics. *Math. Proc. Cambridge Philos. Soc.*, 54(1):60–71.

[48] Nagashima, R., O'Reilly, R. A., and Levy, G. (1969). Kinetics of pharmacologic effects in man: the anticoagulant action of warfarin. *Clin. Pharmacol. Ther.*, 10(1):22–35.

[49] Nanda, S., Moore, H., and Lenhart, S. (2007). Optimal control of treatment in a mathematical model of chronic myelogenous leukemia. *Math. Biosci.*, 210(1):143–156.

[50] Norton, L. (2005). Conceptual and practical implications of breast tissue geometry: toward a more effective, less toxic therapy. *Oncologist*, 10(6):370–381.

[51] Norton, L. and Simon, R. (1977). Growth curve of an experimental solid tumor following radiotherapy. *J. Natl. Cancer Inst.*, 58(6):1735–1741.

[52] Norton, L. and Simon, R. (1986). The Norton-Simon hypothesis revisited. *Cancer Treat. Rep.*, 70(1):163–169.

[53] Nowak, M. A., Michor, F., and Iwasa, Y. (2003). The linear process of somatic evolution. *Proc. Natl. Acad. Sci. U. S. A.*, 100(25):14966–14969.

[54] Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28.

[55] Overgaard, R. V., Jonsson, N., Tornøe, C. W., and Madsen, H. (2005). Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *J. Pharmacokinet. Pharmacodyn.*, 32(1):85–107.

[56] Parra-Guillen, Z. P., Mangas-Sanjuan, V., Garcia-Cremades, M., Troconiz, I. F., Mo, G., Pitou, C., Iversen, P. W., and Wallin, J. E. (2018). Systematic modeling and design evaluation of unperturbed tumor dynamics in xenografts. *J. Pharmacol. Exp. Ther.*, 366(1):96–104.

[57] Perié, L., Hodgkin, P. D., Naik, S. H., Schumacher, T. N., de Boer, R. J., and Duffy, K. R. (2014). Determining lineage pathways from cellular barcoding experiments. *Cell Rep.*, 6(4):617–624.

[58] Ribba, B., Holford, N. H., Magni, P., Trocóniz, I., Gueorguieva, I., and Others (2014). A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. CPT pharmacometrics syst pharmacol 3: e113.

[59] Samara, E. and Granneman, R. (1997). Role of population pharmacokinetics in drug development. *Clin. Pharmacokinet.*, 32(4):294–312.

[60] Schättler, H. and Ledzewicz, U. (2015). *Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods*. Interdisciplinary Applied Mathematics. Springer New York.

[61] Sheiner, L. B. and Beal, S. L. (1980). Evaluation of methods for estimating population pharmacokinetics parameters. i. Michaelis-Menten model: routine clinical pharmacokinetic data. *J. Pharmacokinet. Biopharm.*, 8(6):553–571.

[62] Sheiner, L. B., Rosenberg, B., and Marathe, V. V. (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J. Pharmacokinet. Biopharm.*, 5(5):445–479.

[63] Sheiner, L. B., Rosenberg, B., and Melmon, K. L. (1972). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput. Biomed. Res.*, 5(5):441–459.

[64] Sheiner, L. B., Stanski, D. R., Vozeh, S., Miller, R. D., and Ham, J. (1979). Simultaneous modeling of pharmacokinetics and pharmacodynamics: application to d-tubocurarine. *Clinical Pharmacology & Therapeutics*, 25(3):358–371.

[65] Sheiner, L. B. and Steimer, J. L. (2000). Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu. Rev. Pharmacol. Toxicol.*, 40:67–95.

[66] Sorger, P. K., Allerheiligen, S. R. B., Abernethy, D. R., Altman, R. B., Brouwer, K. L. R., Califano, A., D'Argenio, D. Z., Iyengar, R., Jusko, W. J., Lalonde, R., and Others (2011). Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. In *An NIH white paper by the QSP workshop group*, volume 48.

[67] Stein, S., Zhao, R., Haeno, H., Vivanco, I., and Michor, F. (2018). Mathematical modeling identifies optimum lapatinib dosing schedules for the treatment of glioblastoma patients. *PLoS Comput. Biol.*, 14(1):e1005924.

[68] Stengel, R. F. (1986). *Optimal Control and Estimation*. Courier Corporation.

[69] Swan, G. W. (1990). Role of optimal control theory in cancer chemotherapy. *Math. Biosci.*, 101(2):237–284.

[70] Swan, G. W. and Vincent, T. L. (1977). Optimal control analysis in the chemotherapy of IgG multiple myeloma. *Bull. Math. Biol.*, 39(3):317–337.

[71] Takasu, F. (2011). Lecture 7: Birth-death models 2. *Nara Women's University*, University lecture:1–5.

[72] Terranova, N., Girard, P., Klinkhardt, U., and Munafo, A. (2015). Resistance development: A major piece in the jigsaw puzzle of tumor size modeling. *CPT Pharmacometrics Syst Pharmacol*, 4(6):320–323.

[73] Tornøe, C. W., Overgaard, R. V., Agersø, H., Nielsen, H. A., Madsen, H., and Jonsson, E. N. (2005). Stochastic differential equations in NONMEM®: Implementation, application, and comparison with ordinary differential equations. *Pharm. Res.*, 22(8):1247–1258.

[74] Trocóniz, I. F., Armenteros, S., Planelles, M. V., Benítez, J., Calvo, R., and Domínguez, R. (2000). Pharmacokinetic-Pharmacodynamic modelling of the antipyretic effect of two oral formulations of ibuprofen. *Clin. Pharmacokinet.*, 38(6):505–518.

[75] Trocóniz, I. F., Naukkarinen, T. H., Ruottinen, H. M., Rinne, U. K., Gordin, A., and Karlsson, M. O. (1998). Population pharmacodynamic modeling of levodopa in patients with parkinson's disease receiving entacapone. *Clin. Pharmacol. Ther.*, 64(1):106–116.

[76] Venkatakrishnan, K., Friberg, L. E., Ouellet, D., Mettetal, J. T., Stein, A., Trocóniz, I. F., Bruno, R., Mehrotra, N., Gobburu, J., and Mould, D. R. (2015). Optimizing oncology therapeutics through quantitative translational and clinical pharmacology: challenges and opportunities. *Clin. Pharmacol. Ther.*, 97(1):37–54.

[77] Wagner, J. G. (1968). Kinetics of pharmacologic response. i. proposed relationships between response and drug concentration in the intact animal and man. *J. Theor. Biol.*, 20(2):173–201.

[78] Williams, P. J. and Ette, E. I. (2006). Pharmacometrics: Impacting drug development and pharmacotherapy. In *Pharmacometrics*, pages 1–21.

[79] Woodcock, J. and Woosley, R. (2008). The FDA critical path initiative and its influence on new drug development. *Annu. Rev. Med.*, 59:1–12.

Aim

Pharmacometrics and Systems Pharmacology (PSP) is a continuously evolving science that is being forced to deal with new challenges, as the need to develop multi-scale models using data from different sources, the integration of different modeling strategies coming from Systems Biology and Bioinformatics in order to create more mechanistic models and the lack of tools to perform such integration.

Thus, the primary goal of this thesis was to explore the use of different methodologies and tools that could benefit standard PKPD modelling, highlighting the main advantages and drawbacks of each technique and their contribution to PSP.

To achieve this aim, the specific contributions of the present dissertation include:

- A Boolean modeling framework implemented in an R package called SPIDDOR (Systems Pharmacology for effIcient Drug Development On R) to perform simulations and analyze the results of Boolean networks applied to Systems Pharmacology models. Here not only the specific methodology of logic models is explained but also an open-source tool based on R is presented.
- An optimization procedure based on Optimal Control techniques that aimed to enhance the release characteristics of sustained-release formulations of triptorelin, a gonadotropin-releasing hormone analog used to achieve chemical castration in prostate cancer patients by inhibiting their levels of testosterone.
- A computational framework called ACESO (A Cancer Evolution Simulation Optimizer) to explore the effects of pharmacokinetics and drug interactions in stochastic models describing cancer progression and emergence of drug resistance with the ultimate goal of identifying optimum dosing strategies.
- A semi-mechanistic model describing the time course of several circulating biomarkers in advanced melanoma patients treated with adjuvant high-dose interferon-$\alpha 2b$ used to evaluate the dynamics of the tumor markers as prognostic factors of the overall survival and progression-free survival of the patients. Furthermore, a semi-mechanistic myelosuppression model to evaluate the adverse effects of the therapy is also included in the framework.

1

Advanced Boolean Modeling of Biological Networks Applied to Systems Pharmacology

Advanced Boolean Modeling of Biological Networks Applied to Systems Pharmacology

Itziar Irurzun-Arana, José Martín Pastor, Iñaki F. Trocóniz, José David Gómez-Mantilla

Pharmacometrics & Systems Pharmacology group, Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Navarra, Spain.

IdiSNA, Navarra Institute for Health Research, Pamplona, Navarra, Spain.

Bioinformatics **33(7):1040-1048, April 2017.**

DOI: https://doi.org/10.1371/journal.pcbi.1006087

1^{st} place poster award winner in the Quantitative Systems Pharmacology (QSP) Day 2018 held in MIT Samberg Conference Center (Cambridge, MA, USA). April 2018.

Abstract

Motivation: Literature on complex diseases is abundant but not always quantitative. Many molecular pathways are qualitatively well described but this information cannot be used in traditional quantitative mathematical models employed in drug development. Tools for analysis of discrete networks are useful to capture the available information in the literature but have not been efficiently integrated by the pharmaceutical industry. We propose an expansion of the usual analysis of discrete networks that facilitates the identification/validation of therapeutic targets.

Results: In this article, we propose a methodology to perform Boolean modeling of Systems Biology/Pharmacology networks by using SPIDDOR (Systems Pharmacology for effIcient Drug Development On R) R package. The resulting models can be used to analyze the dynamics of signaling networks associated to diseases to predict the pathogenesis mechanisms and identify potential therapeutic targets.

Availability: The source code is available at https://github.com/SPIDDOR/SPIDDOR.

1.1 Introduction

Computational models are frequently used in the area of biomedicine to interpret, describe or predict dynamic profiles associated to disease progression or drug effects. Among them, the so called population pharmacokinetic/pharmacodynamics (popPKPD) models integrate different type of information, mainly, dosing paradigms and drug exposure, response data, and patient characteristics to account for the time course of drug effects. PopPKPD models are well established in clinical practice and drug development to individualize dosing, identify covariates responsible of inter-patient variability, and dose selection [1, 7, 8]. However, there are several pending challenges in the application of computational models to drug development such as early target identification, choice of best promising drug combinations, understanding resistance development and highlighting patient sub-population sensitive and non-sensitive to a particular therapeutic strategy.

To achieve these goals, popPK/PD models would require a greater mechanistic structure. Nonetheless, mechanistic models require large number of kinetic/dynamic parameters and the task of identifying these parameters is not always possible due to the lack of longitudinal and quantitative data available.

The emergent field of Systems Pharmacology (SP) has the role of bridging System Biology with popPKPD models and it is expected to help in overcoming the bottlenecks highlighted before [4, 13, 11, 20, 26, 30, 40, 42]. SP models can be viewed as networks, which are simplified representations of biological systems in which the components of the system such as genes, proteins or metabolites are represented by nodes and the interactions between them by edges [5, 43]. In general, two different approaches can be used to analyze this type of models: continuous dynamic methods, where the concentrations/amounts of the components are based on differential equations, or discrete dynamic strategies, in which each node can be characterized by only a few discrete states, indicated in contexts where quantitative and longitudinal data are scarce or even not available.

Boolean network models, originally introduced by Kauffman [22, 21], represent the simplest discrete dynamic models. Very briefly, they only assume two discrete states for the nodes of a network, ON or OFF, corresponding to the logic values 1 (active) or 0 (not active, but not necessarily absent). That is why they are known as Boolean or logic models.

A well-designed logic model would be able to generate predictive outcomes given a set of initial conditions. In terms of applications, it would be possible to test how the elimination or overexpression of one or more components of the system affects the final state of the model, which may be useful in the design of combinatorial therapies for a disease or identification of essential components that could be tested as therapeutic targets. It could be also important to detect critical nodes whose perturbation leads to significant functional changes in the system in order to reduce the size of the network by

removing the redundant components. This could be a starting point to try a more quantitative approach.

Currently the application of Boolean analysis to SP is still very limited, contrary to the case of applying dynamic models to continuous or non-continuous data, where there is a battery of tools to help the scientist for model implementation, fitting and evaluation (NONMEM, PsN, Pirana, etc). Consequently, integration of discrete analysis tools in drug development has not been accomplished yet despite its great potential.

Based on these considerations we have developed a framework for an efficient Boolean analysis facilitating (i) model implementation and visualization, (ii) simulation of activation profiles associated with corresponding confidence intervals, (iii) attractor analysis and (IV) a system perturbation and sensitivity analysis. The tools presented in this manuscript consist on a set of comprehensive R scripts to perform discrete dynamic analysis in the context of development therapies for complex diseases.

From a methodological point of view the Boolean analysis presented in this work involves certain novelties. Common Boolean modeling approaches only define direct activation-inhibition relationships between the components of the network. In our models, new types of regulatory interactions have been introduced, the positive and negative modulations, which lead to richer dynamics between the nodes. We also propose a new option to perturb a component of the network emulating a "polymorphism" of a node. Finally, novel approaches were developed for the exploratory analysis of the output of the simulations computed on these models: (i) we incorporate new visualization techniques to evaluate the attractors of the system and the effects of perturbations and (ii) a clustering method is used to group the nodes that lead to similar alterations within the network.

This article guides the reader through the tools developed in our laboratory for an example metabolic network (Figure 1.1) based on a model for immune response to autoantigens [31] and gives a feel of what can be done with its use. The package is called Systems Pharmacology for effIcient Drug Development On R (SPIDDOR). R scripts, help files and vignettes are available in https://github.com/SPIDDOR/SPIDDOR.

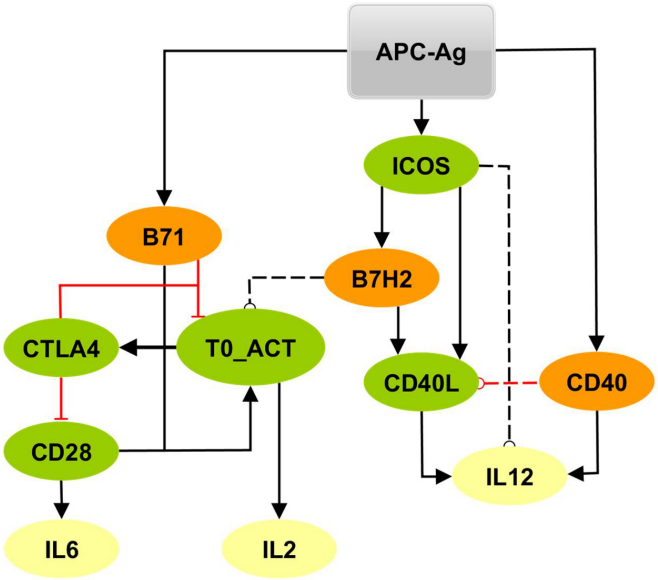


Fig. 1.1 Boolean network example with 12 nodes and 19 regulatory edges made with yEd Graph Editor software. Conceptual nodes (APC-Ag) are presented by a gray rectangle, whereas the molecules of the network are indicated with ellipses. Node colors reflect the nature of the molecules: APC molecules are shown in orange, T cell molecules are in green and interleukins appear in yellow (colored figure online). Arrowheads represent activation, red blunt edges indicate inhibition, black dashed lines imply positive modulations and red dashed lines are negative modulations. The description of the nodes and the Boolean functions is given in Table 1.1.

Table 1.1 Boolean functions of the nodes of the network of Figure 1.1.

Node	Description	Boolean Function	Text file
APC-Ag	Antigen presentation	APC-Ag* = APC-Ag	APC-Ag = APC-Ag
B71	CD80 molecule	B71* = APC-Ag	B71 = APC-Ag
ICOS	Inducible T-cell co-stimulator	ICOS* = APC-Ag	ICOS = APC-Ag
CD40	CD40 molecule	CD40* = APC-Ag	CD40 = APC-Ag
B7H2	ICOS ligand	B7H2* = ICOS	B7H2 = ICOS
CD28	CD28 molecule	CD28* = NOT CTLA4	CD28 = ! CTLA4
CTLA4	Cytotoxic T-lymphocite-associated protein 4	CTLA4*= $\cap_{i=1}^{T0\_ACT^{max}} T0\_ACT^{t-i}$	CTLA4= THR_T0_ACT[3]
CD40L	CD40 ligand	CD40L* = ICOS AND B7H2 AND NOT (CD40 AND CD40L)	CD40L = ICOS & B7H2 & ! (CD40 & CD40L)
T0_ACT	Activated T cell	T0_ACT* = (CD28 AND B71) OR (T0_ACT AND B7H2) AND NOT ($\cap_{i=1}^{MOD} T0\_ACT^{t-i}$ & $\cap_{i=1}^{MOD} B7H2^{t-i}$) AND NOT (CTLA4 AND B71)	T0_ACT = ((CD28 & B71) | (T0_ACT & B7H2) &! (MOD_T0_ACT & MOD_B7H2))) &! (CTLA4 & B71)
IL2	Interleukin 2	IL2* = T0_ACT	IL2 = T0_ACT
IL6	Interleukin 6	IL6* = CD28	IL6 = CD28
IL12	Interleukin 12	IL12* = (CD40 AND CD40L) OR (IL12 AND ICOS) AND NOT ($\cap_{i=1}^{MOD} IL12^{t-i}$ & $\cap_{i=1}^{MOD} ICOS^{t-i}$)	IL12 = (CD40 & CD40L) | (IL12 & ICOS) &! (MOD_IL12 & MOD_ICOS)

The * denotes the future state of a node.

1.2 Methods

Our approach for Boolean modeling biological/pharmacological networks entails the workflow seen in Figure 2.3.

The first step in turning the concepts from literature into a discrete dynamic model is to represent the conceptual model as a directed graph showing the different nodes and the interactions between them. Such networks involve the coordinated interaction of many molecules and stimulus that include genes, proteins, metabolites, cellular states or other conceptual nodes as in Figure 1.1. After defining the components and interactions of the network, the next step is to implement the Boolean transfer functions based on an exhaustive literature research and introduce them in the R environment.

1.2.1 Boolean functions

The state of each node is determined by the state of its regulator nodes (nodes that control its activation/inhibition) based on transition rules known as the Boolean functions (BFs). Depending on the output of the BF, the state of a node can transit from one value to another as the simulation algorithm moves from an iteration to the next. Here, an iteration finishes when all the nodes in the network are updated according to their BFs (in many research works these iterations of the algorithm are referred as time steps but we prefer the term iteration to emphasize that a time step is not necessarily equivalent to a time length). BFs consist on a set of rules specifying how the nodes' states change over time, as a function of the current or past values of its regulator nodes. The main operators of Boolean dynamics are the conjunction AND, the disjunction OR and the negation NOT. Additionally, some convenience operators have been defined. For example, some nodes may need longer activation times of its regulator nodes to be activated. We represent this feature with the $\cap$ notation that can be seen in the BFs of Table 1.1 [37], and we called it threshold operator. The threshold operator requires a duration argument which indicates the number of previous iteration that must be evaluated for a regulator node. In the case of 1.1, it is used to represent that *CTLA-4* molecule is active only if the T cell activation node (*T0_ACT*) is ON for a defined number of iterations indicated by the parameter *T0_ACTmax* (*T0_ACTmax = 3* in our simulations).

Generally, Boolean functions represent simple dynamics of activation and inhibition between nodes. In this work, we present two new possible combinations of Boolean operators that allow us to characterize more precisely some typical processes of biological systems. There are many cases in the literature in which a node A is not able to activate another node B, but A can increase or prolong B expression if B is activated by other signals. We considered this relationship as a positive modulation of node B by node A and we expressed it with the following combination of Boolean operators: *B* = Activators OR (B AND A)*. As can be seen, this regulatory function introduces a self-regulation of the target node. Similarly, if node A cannot directly inhibit node B but it can

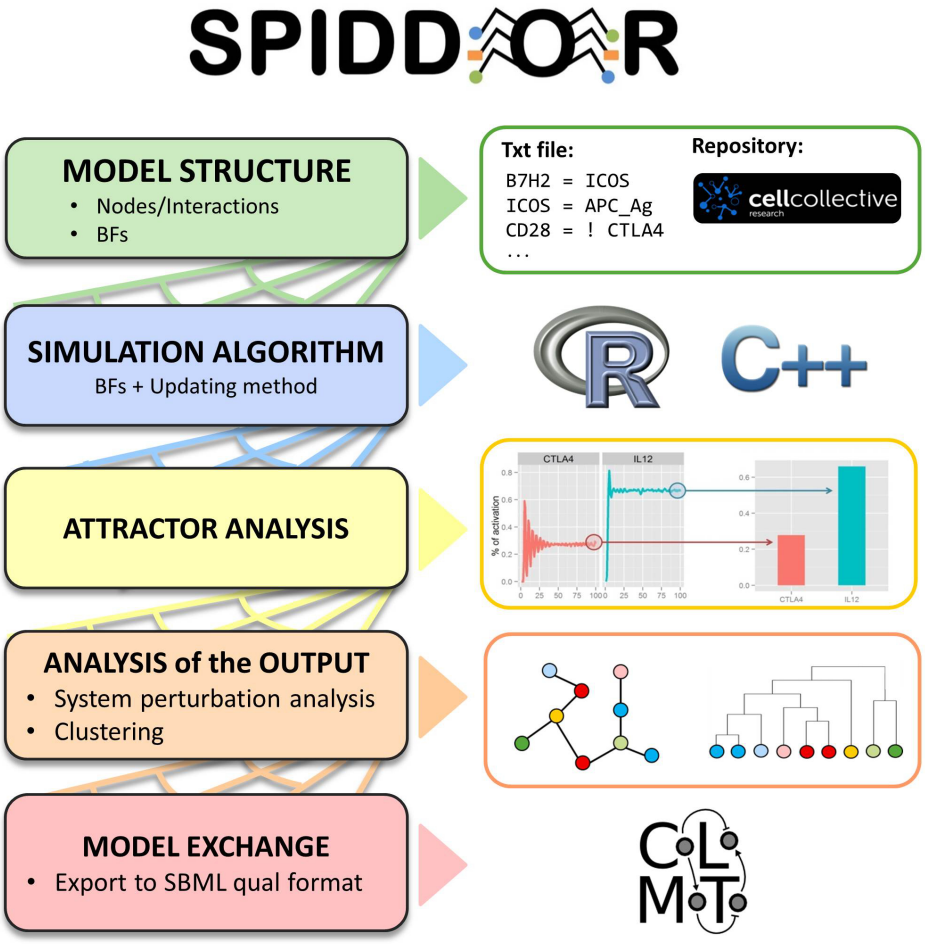


Fig. 1.2 Workflow of the methodology employed by SPIDDOR to perform Boolean modeling of biological networks and the most relevant outputs of each section. First, the model structure is defined in a text file or downloaded from The Cell Collective repository. Second, SPIDDOR reads the BFs from the input files and creates a simulation algorithm in R or C++. Then, the package is able to perform an attractor analysis and introduce perturbations into the model to analyze the output of these networks. Finally, the Boolean networks can be exported to SBML qual format to share models or use other platforms from the CoLoMoTo community.

decrease or shorten its expression it was considered as a negative modulation and we expressed it like *B*=Activators AND NOT (B AND A)*. Furthermore, we have designed these modulatory interactions between nodes to last only a few iterations: *B*= Activators OR ((B AND A) AND NOT* $(\cap_{i=1}^{MOD} B^{t-i}$ *AND* $\cap_{i=1}^{MOD} A^{t-i}))$, with *MOD* argument specifying the maximum number of iterations that the positive modulation will last. All BFs corresponding to the example network are listed in Table 1.1.

BFs are introduced in SPIDDOR by a simple text file written with the appropriate equation semantics (see SPIDDOR vignettes to properly write the input text files) and the system transforms this file into R or C++ code. Another possibility is to load Boolean expressions from a pre-built network from The Cell Collective repository[17], a web-based platform included in the CoLoMoTo (Consortium for Logical Models and Tools) consortium [28].

1.2.2 Nodes updating

The outcome of a Boolean model is also influenced by the chosen updating method, which could be synchronous or asynchronous [15, 37, 33, 41]. The updating method refers to the process of computing the BF of a node to activate or deactivate it in a particular iteration. In a synchronous updating method, the state of the network at each step is determined by the state of the nodes in the prior iteration of the algorithm. In such models, the dynamic trajectory of the network is deterministic, that is, the network will always reach the same state after the same number of iterations. This scheme assumes that all biological processes of the system have similar timescales, which seems quite unrealistic because molecular events are not coordinated in time. A more complex but realistic strategy is the random asynchronous method, where the nodes of the system are updated according to the last update of their regulator nodes, which could be either in the previous or current iteration. In this method, the order in which the nodes update their states is selected randomly during each iteration. This introduces variability into the model, because the same initial conditions can lead to different final states of the network and with different time courses.

Once the BFs are determined, they are implemented in the R environment. A function is written for each BF of the Boolean model, using both the synchronous and asynchronous updating methods. The R script containing the BFs for the example network of Figure 1.1 can be found in the Supplementary Material.

1.2.3 Network evolution in time

We developed a simulation algorithm to calculate the evolution of the network states taking into account the synchronous and asynchronous updating methods, although we recommend the use of the latter as it constitutes a more realistic approach as discussed above. The output of the algorithm is a matrix called *pattern.m* which represents the states of the nodes (ON/OFF) in each step (Figure 1.3A).

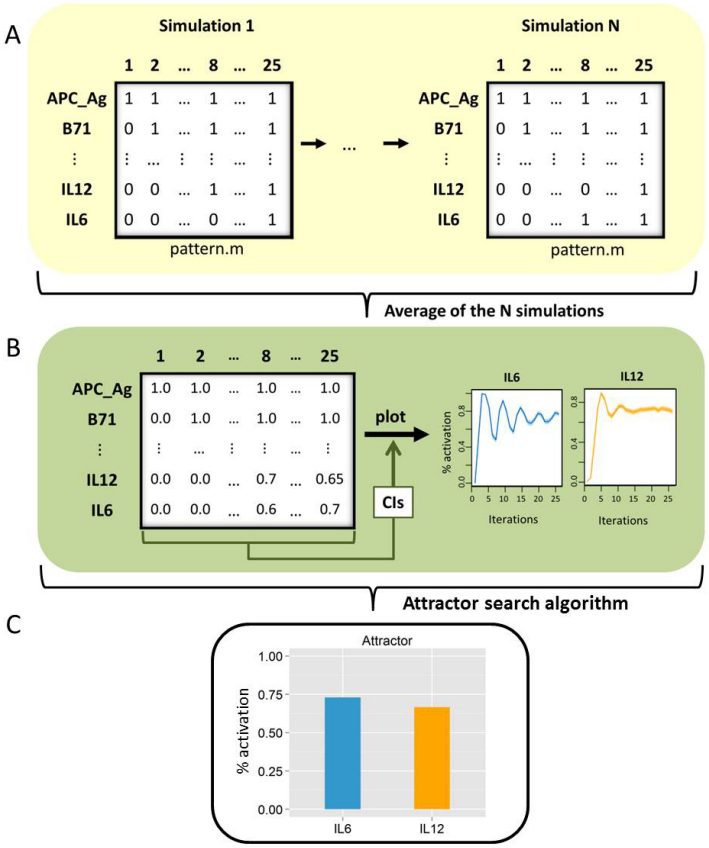


Fig. 1.3 Schematic representation of the steps performed by the asynchronous algorithm. The rows in the matrices correspond to the nodes of the network in Figure 1.1 and the columns to the iterations performed by the algorithm, 25 in this case. (A) The output of the simulation algorithm, the pattern.m matrix. (B) Average of the simulation algorithm results. The average was computed under 2000 (N) simulations in order to calculate the activation profiles of the nodes. (C) Probability of being ON of IL6 and IL12 nodes in the complex attractor found with the asynchronous attractor search algorithm.

It must be stressed that, due to the stochasticity involved in the asynchronous updating scheme, the simulations must be computed repeated times in order to estimate an average of the dynamic trajectory of the network. This allows the calculation of the activation profiles of the nodes for any set of initial conditions. A schematic representation of this process is shown in Figure 1.3B.

To estimate 95% Confidence Intervals (CIs) for the activation profiles of the nodes, we used a method to calculate CIs for proportions by using a binomial distribution described by [29]. For a more detailed description of this method see Supplementary Methods.

1.2.4 Attractor analysis

Starting from an initial condition, Boolean models eventually evolve into a limited set of stable states known as attractors [18]. Once the model has settled onto an attractor, it will remain there for the rest of the simulation. Attractors fall into three groups:

- Fixed-points, which consist of a single attractor state. They are the same for both synchronous and asynchronous update methods because of their time-independence property[33].
- Simple or limit cycles are sets of states in which the system regularly oscillates. These are typical of the synchronous method where each state has only one possible successor state. In our models, however, the states of the nodes in the current iteration not only depend on the states of the nodes in the previous step, but also on prior steps due to the temporal predicates implemented with the threshold operator and the modulators introduced in the system. This produces regular cycles with duplicated states that we called "complex cycles".
- Complex attractors are sets of states in which the system irregularly oscillates due to the randomness involved in asynchronous networks. In these models there is usually more than one possible successor state for each state, so the system does not oscillate in cycles.

Generally, large-scale or highly interconnected networks converge into a complex attractor when an asynchronous updating scheme is used. This oscillatory behavior in Boolean models is due to the presence of negative feedback loops in the network [38, 32, 39]. Attractors in moderate size networks are often linked to cellular steady states, cell cycles, circadian rhythms or to phenotypes [25, 2, 6, 36]. However, it is difficult to make biological inferences from complex attractors as they normally include a high number of stable states that do not oscillate in single cycles.

Our algorithm to identify attractors with the synchronous updating method starts from an initial state and repeatedly performs state transitions until an already visited state is reached. When the synchronous attractors are found (a fixed-point, a simple cycle or a complex cycle) they can be visualized as transition tables where the color inside the table represents the ON/OFF states of the nodes (Supplementary Figure S1). Asynchronous attractor search is more complex as it computes the attractor via exhaustive repetitions of the simulation algorithm. The states in asynchronous attractors do not oscillate cyclically, so they cannot be visualized using transition tables as in the previous case. For this reason, we decided to summarize the information about all the stable states in the attractor by generating the probability that a given node is ON inside the complex attractor. Finally, we visualize these probabilities using bar graphs (Figure 1.3C).

Identification of all the attractors in large-scale asynchronous models is an arduous task due to the computational time required, especially if the attractors are complex because some of their states rarely occur. Moreover, these steady states can change when initial conditions are modified or perturbations are included in the system. We found that the activation probabilities of the nodes in complex attractors almost did not change if the "unusual" states were ignored, suggesting that we could estimate an approximation of the attractor by excluding those rare states from the analysis. This approximation decreases the number of repetitions needed for the asynchronous attractor search algorithm. In addition, for large-scale networks, we recommend coding the simulation algorithm on

C++ to increase speed up to 60-fold. We used the Rcpp R package to communicate R to the C++ algorithm and get the result back to the R environment, allowing its use by normal R users not skilled in C++.

Our main attractor search algorithm is coded to identify the attractor for a given initial condition. We introduced this simplification because we were not interested in testing all the possible initial states, as we typically defined a few possible initial conditions for our networks. However, in some cases there is not enough information to specify the initial condition of a system and sampling of a multitude of initial conditions is necessary. For those interested in this feature, SPIDDOR includes an attractor search algorithm that searches the attractors for networks with less than 20 nodes, as the number of initial conditions to test grow exponentially with the number of nodes. For larger networks, we allow the specification of a subset of nodes (always less than 20) in which all the combinations are to be tested, or the specification of a number of starting states to test (the restriction of maximum 20 nodes limits the initial conditions to test in less than 1,000,000).

Since the hypothetical network used in this article is moderated in size, there is no need of using a parallelized algorithm to reduce the computing time for attractor searching. Even so, this feature is contemplated in our framework and the code for the parallelization using the snowfall library [23] is included in the github repository.

1.2.5 Perturbation analysis

A system perturbation analysis can be performed in order to evaluate which node knockouts or overexpressions lead to significant variations of the network dynamics. A knockout implies the deactivation of a component during all the simulation, whereas an overexpression generates a persistent activation of a node. Another possibility is to overexpress a node but only after its first activation or to activate/deactivate a node for some time. This analysis allows the researcher to model the effects of pharmacological blockades or simulate targeted therapies such as monoclonal antibodies (mAbs).

Our modeling approach also allows the emulation of “polymorphism like” alterations on the components of the network that can result in modifications of their activation patterns. In biology, genetic polymorphisms cause decreased, increased, or absent gene expression or molecular activity by multiple mechanisms. We included these “mutation like” perturbations in which the activity of a node is associated with a probability dependent on the “polymorphism like” conditions. In other words, when a polymorphism was included, we decreased the activity of a node to a lower extent (75%, 50%, 25%...). In this way, when a polymorphism of 50% activity was introduced in a node, this node was activated only 50% of the times in which its regulator nodes were activated.

The activation level of the nodes in normal conditions and when a node was knocked-out or overexpressed were compared in order to analyze how the perturbation of single nodes affected the stable patterns of the rest of the nodes in the network. If the probability of being ON for a node was decreased due to the inclusion of a perturbation, it means that the perturbation caused a lower activation of the component compared to the unperturbed condition. Conversely, if the probability was increased due to a perturbation, it indicates that the perturbation caused a higher activation of the component.

We developed a perturbation analysis algorithm that performs combined synchronous-asynchronous simulations for faster identification of attractors with or without perturbations. First, the program initiates a synchronous attractor search in order to detect whether the network reaches a fixed-point, as this type of attractor is the same in both synchronous and asynchronous algorithms. If this is not the case, we run the asynchronous attractor search to find the complex attractor and the frequency of being ON of each node in these attractors that represent its activation level.

The result of the perturbation analysis is a square matrix in which the number of rows and columns is equal to the number of nodes in the network. It indicates how the knockout/overexpression of the "column node" affected each "row node" (Figure 1.4A). The value in each cell of the matrix corresponds to the probability ratio between the perturbed and the normal conditions. We call to this ratio the *Perturbation Index*(PI) of the nodes. The equation for a given node *i* under a perturbation in *J* is the following: $PI\_J_i = Prob(i)_{Perturbation_J}/Prob(i)_{Normal}$, where *Prob* is the probability of being ON of the node in a given attractor state. Values close to 1 mean that the activity of a node in normal and altered conditions was very similar, and therefore the perturbation had a minor effect on the component.

In order to improve the visualization of this analysis, we transformed the resulting matrix to store only 3 possible values, -1, 0 and 1, as shown in Figure 1.4B. The -1 substitutes the positions where there is a lower activation of a component (value<0.8), the 0 indicates no significant variation between the perturbed and unperturbed conditions, and the 1 represents the locations where there is a higher activation of a node (value>1.25). If a more complex network is being modeled, it is preferable to use more than 3 values to take into account different levels of regulations. The rescaled matrix can be represented using the corrplot package in R in order to visualize the individual values contained in the matrix as colors (Figure 1.4C). In this work, we only performed single node disruptions, altering one-by-one each node from the network, but double or triple perturbations can also be induced in the simulations.

1.2.6 Clustering

Hierarchical clustering methods [14] determine clusters of similar data points based on their distance and build a hierarchical structure on top of them. We applied this method on the results of the perturbation analysis, under the assumption that node alterations that provoke similar effects on the

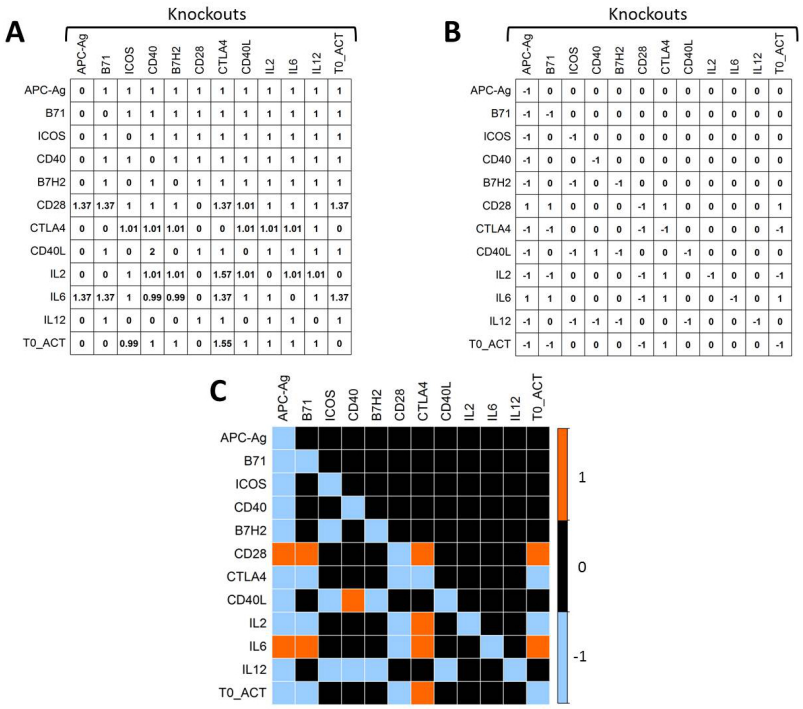


Fig. 1.4 Results arising from a knockout analysis of the network in Figure 1.1 and the subsequent steps to improve its visualization. (A). Numeric matrix with the corresponding *Perturbation indexes* in each cell; (B). Ranking of the values from matrix A; (C). Heatmap of matrix B in which the color indicates if the node knockout entails a lower (blue) or higher (orange) activation of a component compared to an unperturbed simulation.

rest of the nodes of the system will cluster together. Here, we employed the Euclidean metric to determine the distances between each node Perturbation Index and, as merging approach, we used the average-linkage strategy. For example, the distance between a knockout in node A and a knockout in node B would be calculated as follows:

$$d(A,B) = \sqrt{\sum_{i=1}^{n} (PI\_B_i - PI\_A_i)^2}$$

where n is the number of nodes in the network and PI_A and PI_B are the *Perturbation indexes* of the nodes under the knockout in A and B respectively. The results of this exercise are summarized as heatmaps complemented by dendrograms that illustrate the similarity between the perturbations of the system (see Results and Discussion).

1.2.7 Model interoperability

Over the years, different software supporting logical models have been developed, generating different formats to store these models. To address this problem, a novel model exchange format, called SBML qual (Systems Biology Markup Language Qualitative Models) [9], was developed by the CoLoMoTo community [28]. SBML qual is designed for the representation of multivalued qualitative models

of biological networks, thus, enabling models to be shared and used with other platforms and tools without the need of rewriting them in a different format.

We developed a function to export the networks evaluated with SPIDDOR to SBML qual format. We note that SBML does not support networks with temporal operators, so the converter removes these patterns from the nomenclature to store them adequately. In such cases, the output of the simulations performed with SPIDDOR will differ from the results obtained with other platforms because the temporal operators notably change the dynamic evolution of the network.

Models encoded in SBML qual can be submitted to the BioModels database [24] and to The Cell Collective and GINsim [12] software repositories.

1.3 Results and Discussion

In the current work we present the SPIDDOR package which is specifically tailored to the design and analysis of Boolean network models in the area of SP. There are already several software tools and packages available for Boolean modeling of biological systems like BooleanNet [3], BoolNet [27], SimBoolNet [44], ChemChains [16], or GINsim [12]. We note that SPIDDOR differs from other existing methodologies in the following characteristics:

a) Positive and negative modulations: Apart from the basic activation-inhibition interactions, two new regulatory connections have been introduced in the Boolean models, the positive and negative modulations. In the BFs of Table 1.1, IL12 cytokine is positively modulated by the ICOS molecule, meaning that ICOS only activates IL12 if IL12 has already been activated by another regulator node. ICOS does not work as a complete activator because it cannot activate IL12 by itself but it can intensify another activating signal, therefore working as a "sustainer". Similarly, the concept of negative modulator is applied to the CD40 node which does not prevent the activation of node CD40-L by itself but can lessen its expression. Figure 1.5 shows how the activation probabilities of nodes IL12 and CD40L changed when their corresponding logic functions were modified. Both graphs changed when modulation interactions were included compared to simple activations or inhibitions, reflecting the importance of choosing the proper BF for a component. The advantage of incorporating these relationships is that they provide a more semi-quantitative representation of the activity between components, allowing the inclusion of more biologically realistic interactions.

b) Polymorphisms: SP models could be employed to test multiple scenarios as for example the different disease evolution or response to treatment among subject with diverse polymorphisms in a single or various nodes. This perturbation varies a node activity from 0 to 1 and checks the

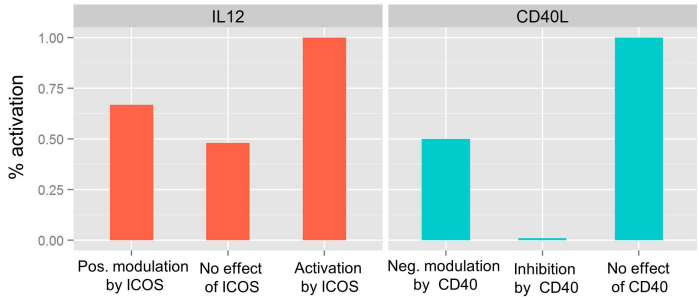


Fig. 1.5 Activation probability of IL12 (left) and CD40L (right) nodes with different Boolean functions. The probability of being ON for IL12 varies when ICOS makes a positive modulation (BF of Table 1.1), a complete activation (IL12 = (CD40 & CD40L) | ICOS) or has no effect on IL12 (IL12 = CD40 & CD40L). On the other hand, the output of CD40L changes when we introduce a negative modulation by CD40 node (CD40L= ICOS & B7H2 &! (CD40 & CD40L)), a complete inhibition by CD40 (CD40L= ICOS & B7H2 &! CD40L) or when CD40 has no effect on CD40L (CD40L= ICOS & B7H2).

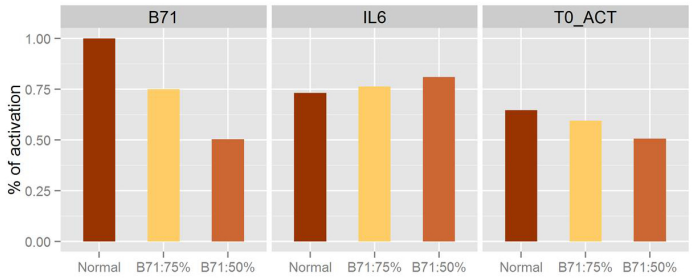


Fig. 1.6 Relative expression profiles of B71, IL6 and T0_ACT with different levels of polymorphisms acting on B71. A polymorphism was simulated on B71 node to reduce a 25% and a 50% its activity. This perturbation increased the levels of IL6 expression and decreased the activation of T cell T0_ACT node.

effect of these variations on a desired outcome. In Figure 1.6 it is shown how polymorphisms acting on the activity of B71 node decreased the activation levels of T0_ACT and increased the expression of IL6 compared to normal response. This type of analysis can be used to evaluate gene mutations that are linked to a particular disease, and test which polymorphisms could trigger similar molecular alterations as the ones reported for the disease. This perturbation analysis is complementary to the introduction of node knockouts or overexpressions which are not realistic representations of human physiopathology. Furthermore, the introduction of a node knockout could provoke a total blockage of one or several pathways hindering the analysis of less severe perturbation or complementary perturbations in other nodes. A similar analysis could be used to explore the effect of target engagement on drug treatment. For example, in Supplementary Figure S2, when a mAb is used as a therapeutic agent, it can be studied what is the required level of target inhibition for an anti-Icos mAb to achieve a reduction of 50% on IL12 expression. Similarly, if a polymorphism is introduced on an input node by setting it to a specific average level of activation, it is possible to explore different background noise levels on the system and evaluate the impact of environmental fluctuations [10].

c) Visualization of attractor states: Some of the mentioned discrete modeling tools like Boolnet have functions to visualize the complex attractors as interconnected graphs representing state transitions inside the attractor. However, when the number of nodes in the network is high, these graphs are extremely difficult to analyze and may not provide meaningful information for the scientists who are not familiar with such discrete outputs. For this reason, we improved their visualization by representing the activation probability of the nodes using bar graphs. For the attractor analysis of the example network we simulated the evolution of the system under a continuous antigen presentation (APC-Ag = 1) in synchronous and asynchronous mode. Under the synchronous updating method, we found a “complex cycle” composed of 28 states (Supplementary Figure S1). The asynchronous attractor search algorithm with 1000 simulation steps and repeated 16 times found a complex attractor composed of 84 states whose activation probabilities are summarized in Supplementary Table S1. In Figure 1.7, we simulated network

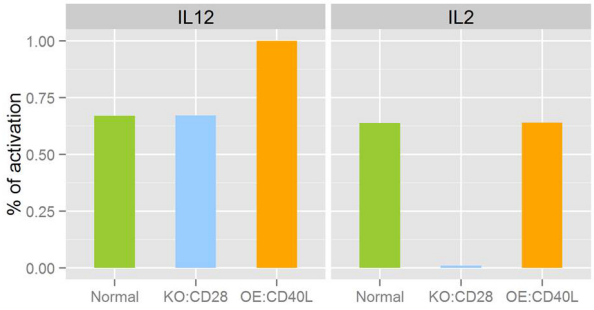


Fig. 1.7 Activation levels of IL2 and IL12 with different perturbations of the system. Two different perturbations were introduced in the model, knock-out of node CD28 (KO:CD28) and over-expression of node CD40L (OE:CD40L), in order to see how the probability of being ON of IL2 and IL12 change.

perturbations by introducing a knockout on CD28 molecule and an overexpression of CD40L and analyze how these alterations affect to IL2 and IL12 activation probabilities.

d) Visualization of perturbation analysis: Several tools were developed for the exploratory analysis of the network output to evaluate many nodes perturbations at the same time on the attractors of the system and to cluster them according to the effects that they provoke. In Figure 1.8A it is shown how a knockout on APC-Ag node modifies the activation probability of all the nodes in the network (shown in orange and blue) as it is the input node of the system, while a knockout on ICOS molecule only downregulates B7H2, CD40L and IL12 components (shown in blue). This is quite easy to infer by observing the structure of the network in Figure 1, but in larger systems the effects of the manipulations are not so easily deduced. The result of the hierarchical clustering calculation is displayed as a dendrogram in the top of these heatmaps (Figure 1.8). For example, the effects of ICOS and B7H2 knockouts in the system are very similar, so they are clustered together in the dendrogram of Figure 1.8A.

The dynamic perturbation analysis is a technique used to identify critical nodes and facilitate network validation. In this type of modeling frameworks, it is possible to emulate a disease on the biological network under study by changing the initial conditions of the computer simulations. Thanks to the dynamic perturbation method it is possible to test which perturbations can revert the disease condition [34]. Such results could be used to prioritize which of the knockouts or constitutive activations should be studied first in wet bench experiments.

Another practical use of the visualization technique explained above is the possibility of performing a sensitivity analysis of the network to discover which nodes have a higher impact on other components of the system. In the matrices of Figure 1.4, nodes ICOS, CD40, CD40L and B7H2 have a higher influence on IL12 interleukin, as their perturbation lead to a significant downregulation of the molecule. If the interest lies mainly in the response of this component,

a network reduction could be applied by removing the components that do not regulate IL12. This ability is important to reduce the size of complex SP networks. This sensitivity analysis can be complemented by the use of the polymorphism tool introduced before to identify sources of interindividual variability by highlighting the nodes which polymorphisms are more likely to provoke large changes in specific outputs.

The immune network presented in this work is an illustrative example used to describe the new methodologies which application we consider useful in the SP field. The results obtained from our simulations should not be considered as a full representation of the immune response because many immunological components have been left apart for simplification purposes.

Despite the advantages of Boolean networks and the methodologies presented in this work, some limitations need to be considered. First, it is important to realize that manually building a biological network may be time-consuming and (inevitably) subjective as BFs are established following the researcher criteria. Some tools exist to infer networks automatically from experimental data [35] but different algorithms lead to different networks while different networks are generally deduced from different datasets, therefore, it is also subjective which algorithm and dataset to use.

A main limitation lies in the reliability of these models. In this project, we tried to reproduce the experimental observations discussed on the research articles used to build the network. The heatmaps created with our framework are useful for this task. For example, in the heatmap of Figure 1.8A, we

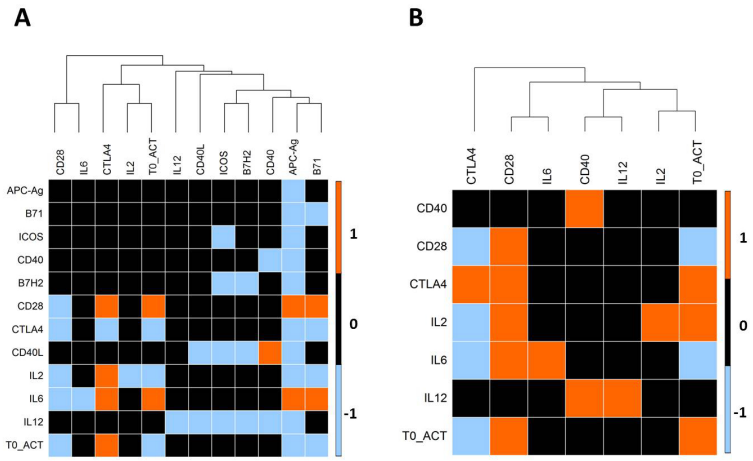


Fig. 1.8 Hierarchical clustering of the perturbations induced on the nodes of the network in Figure 2.1. Heatmaps indicate the effect of single perturbations (knockouts on the left and overexpressions on the right) on the nodes of the network. The perturbations that lead to a higher activation of the nodes compared to an unperturbed situation are represented in orange while a lower activation of the nodes is indicated in blue.

can see how a knockout in CD28 leads to a lower activation of IL2, which is consistent with the results found in the literature [19]. However, this could be a complex task when there is a lack of information about the nodes under study. We are currently working on new possible validation methods based on microarray or RNA-seq analysis, but further work needs to be done.

1.4 Conclusion

Computational models have been increasingly used to support drug development and are widely accepted by scientific community and even for regulatory purposes. A key challenge when using these powerful approaches is to match the right model with the right questions in a particular research context. Although Boolean networks cannot be used for precise estimations such as drug dosing in pediatric or renal impairment population, they are useful to gain insight into the qualitative behavior of a system under study. This is especially relevant for large scale systems in which a detailed kinetic characterization of the system is not feasible due to data restrictions or limited knowledge. More precise quantitative models require exponentially more complex and quality data, and sometimes, acquisition of such data could be restricted by technical constraints, as is the case of immunology in which there are not available techniques for the continuous *in-vivo* measurement of cells subpopulations and cytokines in different tissues. Therefore, it is mandatory to get the best use of the available knowledge in each stage of development, for which it is essential to explore the full potential of tools like Boolean networks. We consider that the methodologies presented in this work can potentiate the use of Boolean networks in SP by introducing versatile tools to enrich the analysis of these systems.

References

[1] Admiraal, R., van Kesteren, C., Boelens, J. J., Bredius, R. G. M., Tibboel, D., and Knibbe, C. A. J. (2014). Towards evidence-based dosing regimens in children on the basis of population pharmacokinetic pharmacodynamic modelling. *Arch. Dis. Child.*, **99**(3), 267–272.

[2] Akman, O. E., Watterson, S., Parton, A., Binns, N., Millar, A. J., and Ghazal, P. (2012). Digital clocks: simple boolean models can quantitatively describe circadian systems. *J. R. Soc. Interface*, **9**(74), 2365–2382.

[3] Albert, I., István, A., Juilee, T., Song, L., Ranran, Z., and Réka, A. (2008). Boolean network simulations for life scientists. *Source Code Biol. Med.*, **3**(1), 16.

[4] Bai, J. P. F., Fontana, R. J., Price, N. D., and Sangar, V. (2014). Systems pharmacology modeling: an approach to improving drug safety. *Biopharm. Drug Dispos.*, **35**(1), 1–14.

[5] Berger, S. I. and Iyengar, R. (2009). Network analyses in systems pharmacology. *Bioinformatics*, **25**(19), 2466–2472.

[6] Bilsland, A. E., Stevenson, K., Liu, Y., Hoare, S., Cairney, C. J., Roffey, J., and Keith, W. N. (2014). Mathematical model of a telomerase transcriptional regulatory network developed by cell-based screening: analysis of inhibitor effects and telomerase expression mechanisms. *PLoS Comput. Biol.*, **10**(2), e1003448.

[7] Borrat, X., Trocóniz, I. F., Valencia, J. F., Rivadulla, S., Sendino, O., Llach, J., Muñoz, J., Castellví-Bel, S., Jospin, M., Jensen, E. W., Castells, A., and Gambús, P. L. (2013). Modeling the influence of the A118G polymorphism in the OPRM1 gene and of noxious stimulation on the synergistic relation between propofol and remifentanil: sedation and analgesia in endoscopic procedures. *Anesthesiology*, **118**(6), 1395–1407.

[8] Buil-Bruna, N., Dehez, M., Manon, A., Nguyen, T. X. Q., and Trocóniz, I. F. (2016). Establishing the quantitative relationship between lanreotide autogel®, chromogranin a, and Progression-Free survival in patients with nonfunctioning gastroenteropancreatic neuroendocrine tumors. *AAPS J.*

[9] Chaouiya, C., Bérenguier, D., Keating, S. M., Naldi, A., van Iersel, M. P., Rodriguez, N., Dräger, A., Büchel, F., Cokelaer, T., Kowal, B., Wicks, B., Gonçalves, E., Dorier, J., Page, M., Monteiro, P. T., von Kamp, A., Xenarios, I., de Jong, H., Hucka, M., Klamt, S., Thieffry, D., Le Novère, N., Saez-Rodriguez, J., and Helikar, T. (2013). SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC Syst. Biol.*, **7**, 135.

[10] Domedel-Puig, N., Rué, P., Pons, A. J., and García-Ojalvo, J. (2011). Information routing driven by background chatter in a signaling network. *PLoS Comput. Biol.*, **7**(12), e1002297.

[11] Geerts, H., Roberts, P., and Spiros, A. (2015). Assessing the synergy between cholinomimetics and memantine as augmentation therapy in cognitive impairment in schizophrenia. a virtual human patient trial using quantitative systems pharmacology. *Front. Pharmacol.*, **6**, 198.

[12] Gonzalez, A. G., Naldi, A., Sánchez, L., Thieffry, D., and Chaouiya, C. (2006). GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Biosystems.*, **84**(2), 91–100.

[13] Goryanin, I. I. and Goryachev, A. B. (2011). *Advances in Systems Biology*. Springer Science & Business Media.

[14] Hartigan, J. A. (1975). *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition.

[15] Harvey, I. and Bossomaier, T. (1997). Time out of joint: Attractors in asynchronous random boolean networks. In *Proceedings of the Fourth European Conference on Artificial Life*, pages 67–75.

[16] Helikar, T. and Rogers, J. A. (2009). ChemChains: a platform for simulation and analysis of biochemical networks aimed to laboratory scientists. *BMC Syst. Biol.*, **3**, 58.

[17] Helikar, T., Kowal, B., McClenathan, S., Bruckner, M., Rowley, T., Madrahimov, A., Wicks, B., Shrestha, M., Limbu, K., and Rogers, J. A. (2012). The cell collective: toward an open and collaborative approach to systems biology. *BMC Syst. Biol.*, **6**, 96.

[18] Hopfensitz, M., Müssel, C., Maucher, M., and Kestler, H. A. (2012). Attractors in boolean networks: a tutorial. *Comput. Stat.*, **28**(1), 19–36.

[19] Howland, K. C., Ausubel, L. J., London, C. A., and Abbas, A. K. (2000). The roles of CD28 and CD40 ligand in T cell activation and tolerance. *J. Immunol.*, **164**(9), 4465–4470.

[20] Iyengar, R., Zhao, S., Chung, S.-W., Mager, D. E., and Gallo, J. M. (2012). Merging systems biology with pharmacodynamics. *Sci. Transl. Med.*, **4**(126), 126ps7.

[21] Kauffman, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, **22**(3), 437–467.

[22] Kauffman, S. A. (1993). *The origins of order: Self organization and selection in evolution*. Oxford University Press, USA.

[23] Knaus, J. (2013). snowfall: Easier cluster computing (based on snow).

[24] Le Novere, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., and Others (2006). BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.*, **34**(suppl 1), D689–D691.

[25] Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. U. S. A.*, **101**(14), 4781–4786.

[26] Lu, Y., Griffen, S. C., Boulton, D. W., and Leil, T. A. (2014). Use of systems pharmacology modeling to elucidate the operating characteristics of SGLT1 and SGLT2 in renal glucose reabsorption in humans. *Front. Pharmacol.*, **5**, 274.

[27] Müssel, C., Hopfensitz, M., and Kestler, H. A. (2010). BoolNet–an R package for generation, reconstruction and analysis of boolean networks. *Bioinformatics*, **26**(10), 1378–1380.

[28] Naldi, A., Monteiro, P. T., Müssel, C., Consortium for Logical Models and Tools, Kestler, H. A., Thieffry, D., Xenarios, I., Saez-Rodriguez, J., Helikar, T., and Chaouiya, C. (2015). Cooperative development of logical modelling standards and tools with CoLoMoTo. *Bioinformatics*, **31**(7), 1154–1159.

[29] Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat. Med.*, **17**(8), 857–872.

[30] Palmér, R., Nyman, E., Penney, M., Marley, A., Cedersund, G., and Agoram, B. (2014). Effects of il-1β–blocking therapies in type 2 diabetes mellitus: A quantitative systems pharmacology modeling approach to explore underlying mechanisms. *CPT Pharmacometrics Syst. Pharmacol.*, **3**(6), e118.

[31] Ruiz-Cerdá, M. L., Irurzun-Arana, I., González-Garcia, I., Hu, C., Zhou, H., Vermeulen, A., Trocóniz, I. F., and Gómez-Mantilla, J. D. (2016). Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. *Eur. J. Pharm. Sci.*

[32] Saadatpour, A. and Albert, R. (2013). Boolean modeling of biological regulatory networks: A methodology tutorial. *Methods*, **62**(1), 3–12.

[33] Saadatpour, A., Albert, I., and Albert, R. (2010). Attractor analysis of asynchronous boolean models of signal transduction networks. *J. Theor. Biol.*, **266**(4), 641–656.

[34] Saadatpour, A., Wang, R.-S., Liao, A., Liu, X., Loughran, T. P., Albert, I., and Albert, R. (2011). Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia.

[35] Scutari, M. (2010). Learning bayesian networks with the bnlearn R package. *J. Stat. Softw.*, **35**(1), 1–22.

[36] Sun, Z., Jin, X., Albert, R., and Assmann, S. M. (2014). Multi-level modeling of light-induced stomatal opening offers new insights into its regulation by drought. *PLoS Comput. Biol.*, **10**(11), e1003930.

[37] Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, E. T., and Albert, R. (2007). Modeling systems-level regulation of host immune responses. *PLoS Comput. Biol.*, **3**(6), e109.

[38] Thomas, R. and d'Ari, R. (1990). *Biological feedback*. CRC press.

[39] Thomas, R., Thieffry, D., and Kaufman, M. (1995). Dynamical behaviour of biological regulatory networks—i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bltn Mathcal Biology*, **57**(2), 247–276.

[40] van der Graaf, P. H. and Benson, N. (2011). Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm. Res.*, **28**(7), 1460–1464.

[41] Wang, R.-S., Saadatpour, A., and Albert, R. (2012). Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.*, **9**(5), 055001.

[42] Wang, Y., Guo, Z., Chen, X., Zhang, W., Lu, A., and Wang, Y. (2015). Multi-scale modeling of cell survival and death mediated by the p53 network: a systems pharmacology framework. *Mol. Biosyst.*

[43] Zhao, S. and Iyengar, R. (2012). Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu. Rev. Pharmacol. Toxicol.*, **52**, 505–521.

[44] Zheng, J., Zhang, D., Przytycki, P. F., Zielinski, R., Capala, J., and Przytycka, T. M. (2010). SimBoolNet—a cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics*, **26**(1), 141–142.

Supplementary Material

Confidence Intervals

95% Confidence Intervals (CIs) for the activation profiles of the nodes are estimated by using a binomial distribution method described by (Newcombe, 1998). This is a discrete probability distribution used to obtain the probability of observing a number of x successes in n independent yes/no experiments. In our case, the successes are the ON states of the nodes. We used the normal approximation of the binomial distribution theorem:

$$CI = p \pm z\sqrt{\frac{p(1-p)}{n}}$$

where CI represents the limits of the confidence interval, p is the activation probability of a node at a particular iteration (between 0-1), n is the sample size (2000 in Figure 3), and z is the value corresponding to a level of confidence of 95% and it is equal to 1.959964.

CIs are used to indicate how much the activation probabilities are affected by the included randomness in the asynchronous method. Explicitly, the CI shows how much the activation probabilities can vary if the network evolutions are simulated with a particular sample size (2000 for activation profiles in Figure 3). Initially, we performed 2000 repetitions of the simulation algorithm and repeated the procedure 2500 times in order to calculate the 97.5 and 2.5 percentiles of the nodes and plot those values in the graphs. However, we realized that using the normal approximation of the binomial distribution theorem gives almost identical results without the need of computing additional simulations.

Supplementary Code

Supplementary code is available at *Bioinformatics* online.

Supplementary Tables

Table S1.8 Activation probabilities of the nodes in the complex attractor of Figure 1.

Node	% of activation
APC-Ag	1.00%
B71	1.00%
ICOS	1.00%
CD40	1.00%
B7H2	1.00%
CD28	0.73%
CTLA4	0.28%
CD40L	0.50%
T0_ACT	0.64%
IL2	0.73%
IL6	0.67%
IL12	0.64%

Supplementary Figures

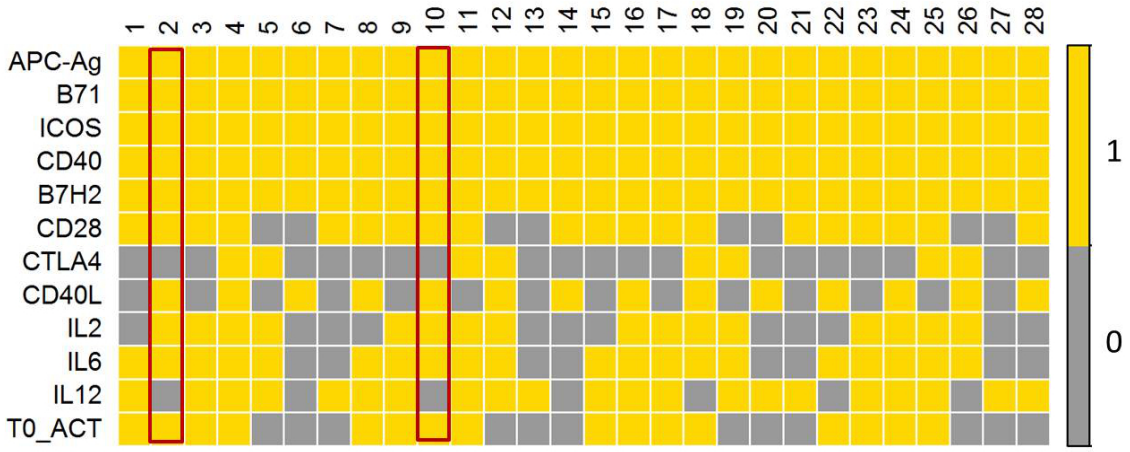


Fig. S1.1 Complex cycle composed of 28 states. The columns of the colored table represent consecutive states of the cycle. Yellow cells denote the activation of a node, whereas grey cells denote deactivation. 10 of the 28 states of the cycle are repeated attractor states (2 of them marked with a red rectangle).

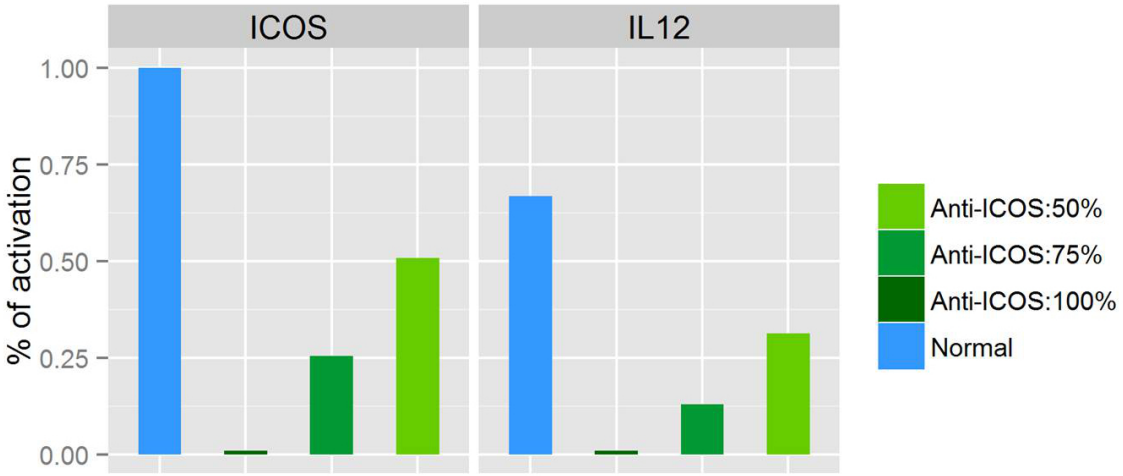


Fig. S1.2 Effect of different levels of anti-ICOS treatment. Different levels of inhibition of ICOS node were simulated (50%, 75% and 100% inhibition) in order to test its effect on IL12. Higher levels of inhibition on ICOS led to a lower activation of IL12 too.

2

Optimal dynamic control approach in a multi-objective therapeutic scenario: Application to drug delivery in the treatment of prostate cancer

Optimal dynamic control approach in a multi-objective therapeutic scenario: Application to drug delivery in the treatment of prostate cancer

Itziar Irurzun-Arana[1,2], Alvaro Janda[3], Sergio Ardanza-Trevijano[3], Iñaki F. Trocóniz [1,2]

1 Pharmacometrics & Systems Pharmacology group, Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Navarra, Spain.

2 IdiSNA, Navarra Institute for Health Research, Pamplona, Navarra, Spain.

3 Department of Physics and Applied Mathematics, University of Navarra, Pamplona, Navarra, Spain.

***PLoS Computational Biology* 14 (4): e1006087, April 2018.**
DOI: https://doi.org/10.1371/journal.pcbi.1006087

Mathematical & Computational Science Special Interest Group (SIG) Award. American Conference on Pharmacometrics 9 (ACoP9) (San Diego, CA, USA). October 2018.

Abstract

Numerous problems encountered in computational biology can be formulated as optimization problems. In this context, optimization of drug release characteristics or dosing schedules for anticancer agents has become a prominent area not only for the development of new drugs, but also for established drugs. However, in complex systems, optimization of drug exposure is not a trivial task and cannot be efficiently addressed through trial-error simulation exercises. Finding a solution to those problems is a challenging task which requires more advanced strategies like optimal control theory. In this work, we perform an optimal control analysis on a previously developed computational model for the testosterone effects of triptorelin in prostate cancer patients with the goal of finding optimal drug-release characteristics. We demonstrate how numerical control optimization of non-linear models can be used to find better therapeutic approaches in order to improve the final outcome of the patients.

Author summary

Mathematical models of the disease processes are widely used in computational biology to quantitatively describe the time course of disease progression and are often linked to pharmacokinetic–pharmacodynamic models in order to evaluate the effect of drug treatment on disease. Once the models are built from observed information and/or literature data, they can predict the dynamics of the system under different conditions through computer simulations. However, simulation exercises are not always effective to obtain the desired objectives due to the complexity of these systems. In this work, we optimized the release characteristics of a synthetic gonadotropin-releasing hormone analog used to induce chemical castration by inhibiting the testosterone levels in prostate cancer patients. The therapeutic goals to achieve were to minimize the initial flare up of testosterone levels and the time to reach testosterone values below castration limit, while maximizing the castration period of the patients. Our methodology, based on control theory, introduces a manipulable variable into the system's equations to drive the model towards the established goals. We demonstrated how drug-release properties can be improved with the implementation of optimal control strategies to enhance the outcome of cancer patients. These methods are extrapolable to other problems encountered in the field.

2.1 Introduction

Optimizing delivery systems targeting constant levels of drug concentration represents always a challenge for chronic diseases requiring continuous treatment and especially in those cases where the relationship between drug exposure (represented generally as levels of drug concentration plasma measured longitudinally) and pharmacological response is complex and non-linear. The management of prostate cancer with sustained release formulations of triptorelin (TRP) injected every 3-6 months represents a good example [1]. For the case of the hormone-sensitive prostate tumors the therapeutic goal of any pharmacology treatment is to maintain as longer as possible the levels of testosterone (TST) below the castration limit (CT) which is set to the plasma concentration value of 0.5 ng/mL [2].

In recent past, we have developed a mechanistic computational model for the TST effects of the agonist TRP in prostate cancer patients using longitudinal pharmacokinetic (PK; drug concentration in plasma) and pharmacodynamics (PD; TST concentrations in plasma) data obtained from several clinical trials testing the efficacy of different sustained-release formulations (SR) [3]. Briefly, TRP exerts its action by increasing the fraction of activated receptors and therefore stimulating the production of TST. However, the prolonged exposure of TRP causes receptor down-regulation, resulting in a reduced synthesis of TST. The typical TST vs time profile after a single injection of TRP is represented in Fig 2.1. The schematic representation of the PKPD model developed for TST effects of TRP, excluding the absorption compartments of the original model, and the estimates of model parameters are shown in Fig 2.2.

As highlighted in Fig 2.1 there are three critical aspects to be taken into consideration at the time to develop an innovative delivery system of TRP for the treatment of prostate cancer: initial flare up, time to reach CT, and castration period. Ideally, such new formulation should release TRP at a rate eliciting levels of concentration in plasma minimizing both the initial flare up and the time to reach CT, as well as maximizing the castration period. Specifically, limitation in the TST flare-up (TST_{max}) to 50% increase with respect to baseline, minimize time to castration after first injection (t_{cast}) to values below 3 weeks, and extend the castration time after injection (t_{effect}) for at least 9 months.

Given the complex relationship between concentrations of TRP in plasma and response as represented in Fig 2.1 and Fig 2.2, together with the requisite of maintaining the TST profiles within the constraints mentioned above, optimization of the rate of drug release is not a trivial task and cannot be efficiently addressed through an extensive trial & error simulation exercise.

In the current work we aimed to optimize the release profile of TRP from SR formulations matching the multi-objective therapeutic needs applying optimal control methodology [4]. The rationale behind the decision of focusing on the release process is based on the assumption that once the drug is absorbed and reaches systemic circulation (represented as part of the central compartment in Fig 2.2)

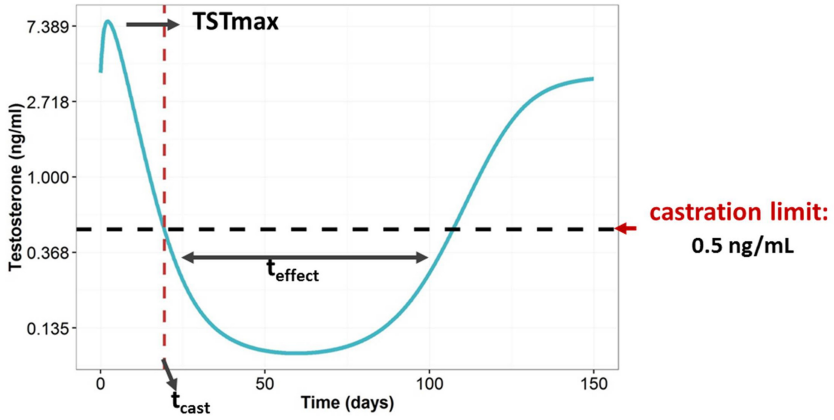
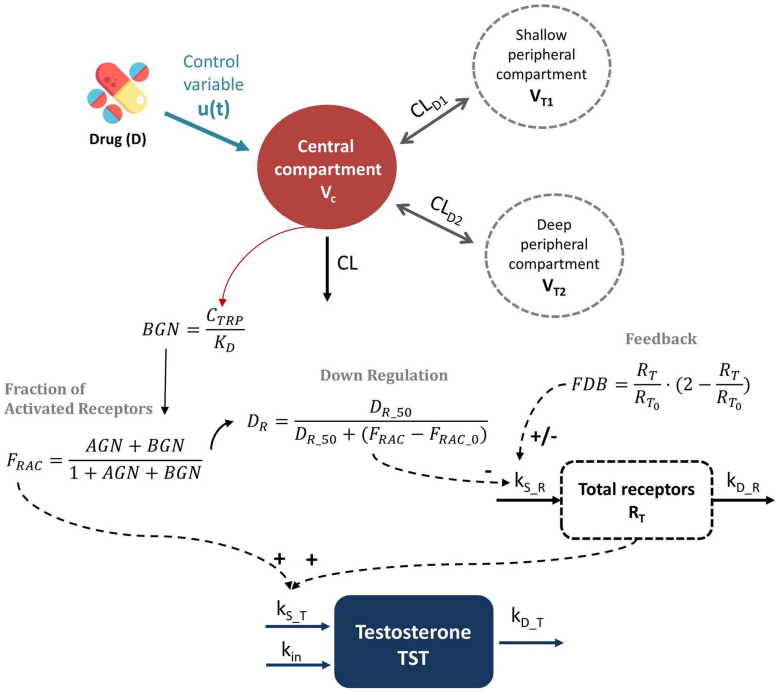


Fig. 2.1 Typical testosterone profile after administration of triptorelin. TST_{max} refers to the maximal testosterone concentrations, t_{cast} indicates the time where testosterone levels fall below 0.5ng/ml (castration limit of prostate cancer patients, marked with an horizontal dashed line in the figure) and t_{effect} indicates the castration period of the patients.

it follows the same distribution and elimination characteristics regardless the type of formulation administered. The same is assumed with respect to the TST response, the rate of synthesis and degradation of TST and receptors, the dynamics of receptor occupation, and the down-regulation process. These mechanisms are independent from the absorption properties of the drug.

Despite we focused on a specific case, the workflow and methodology used can be readily translate to other therapeutic areas and scenarios such as dosing schedule optimization and personalized treatments.



Parameter	Estimate	BSV(%)
$CL(L \cdot day^{-1})$	274.3	10
$V_c(L)$	8.1	37
$V_{T1}(L)$	12	-
$V_{T2}(L)$	33.8	-
$TST_0(ng/mL)$	3.98	35.07
$K_D(ng/mL)$	0.931	-
$D_{R\_50}(unitless)$	0.024	31.36
$k_{S\_R}(day^{-1})$	0.185	32.09
$k_{in}(ng \cdot mL^{-1}day^{-1})$	0.041	35.77
$k_{D\_T}(day^{-1})$	0.55	-
AGN(unitless)	0.31	-

Fig. 2.2 Schematic representation of the state variables and control input for the pharmakinetic-pharmacodynamic model of the testosterone effects of Triptorelin (left) and model parameter estimates (right). C_{TRP}, serum concentrations of Triptorelin; CL, apparent total clearance; V_c, V_{T1}, and V_{T2}, apparent volumes of distribution of the central, shallow, and deep peripheral compartments, respectively; CL_{D1} and CL_{D2}, distribution clearances between the central and peripheral compartments; TST_0, baseline testosterone level; K_D, receptor equilibrium dissociation constant of triptorelin; D_R, down-regulation process; D_{R_50}, the value that elicits a 50% maximal reduction in k_{S_R} for a given amount of total receptors; k_{S_R}, zero-order rate constants of receptor synthesis; k_{D_T}, first-order rate constants of testosterone degradation; k_{in}, zero-order rate production of testosterone independent from gonadotropins; AGN, ratio between the endogenous agonist concentration and its receptor equilibrium dissociation constant; FDB, feedback.

2.2 Materials and methods

Havind defined the therapeutic goals of the project, the analysis was divided in several steps: i) a population of virtual subjects were generated in order to have a representative population of the study; ii) the optimal TST profiles for each virtual patient were derived by means of optimal control methods; and finally iii) the empirical absorption profiles obtained in step ii were characterized using parametric models to assist biopharmaceutics at the time to develop and evaluate new SR formulations of Triptorelin. A schematic representation of the workflow is given in Fig 2.3.

2.2.1 Generation of a virtual patient population

Values listed in the table inserted in Fig 2.2 include estimates of typical population parameters and between-subject variability (represented by BSV in the table and hereafter) obtained from [3] for a set of model parameters. In order to obtain the population of virtual patients, parameters were modelled as $P_i = P_{pop} \times e^{\eta_{i\_P}}$, where P_i and P_{pop} represent the i^{th} individual and typical population values of the P parameter, respectively, and η_{i_P} corresponds the deviation of P_i with respect the typical value P_{pop}; the set of individual η_{i_P} forms a random variable with mean value of 0 and variance ω_P^2 following a normal distribution, whereas the distribution of individual parameters is log-normal. The magnitude of ω_P^2 reflects the BSV associated to a specific model parameter, which in Fig 2.2 is expressed as coefficient of variation (CV%).

One thousand set of disposition (clearance and volume of distribution in the central compartment, represented as CL and V_c respectively), pharmacodynamics (receptor equilibrium dissociation constant of triptorelin (K_D)) and system (baseline TST levels (TST_0), zero-order rate of TST production independent from gonadotropins (k_{in}), zero-order rate constants of receptor synthesis (k_{S_R}) and the value that elicits a 50% maximal reduction in k_{S_R} for a given amount of total receptors (D_{R_50})) related parameters were generated using the typical population estimates and corresponding marginal distributions reported in the table of Fig 2.2. The parameter values for the virtual population were generated with NONMEM 7.2 [5].

2.2.2 Optimal control

An optimal control problem is a dynamic optimization problem in which the state of a system is linked in time to the application of a control function $u(t)$, which drives the system towards a desirable outcome by minimizing a cost function $J(u)$ subject to operating constraints [6, 4]. In other words, the control variable $u(t)$ forces a system to have an optimal performance. The concrete control strategy will depend upon the criterion used to decide what is meant by "optimal"; in the current case $TST_{max} < 1.5 \cdot TST_0$ ng/ml, minimize $t_{cast} \leq 3$ weeks, and maximize $t_{effect} \geq 9$ months.

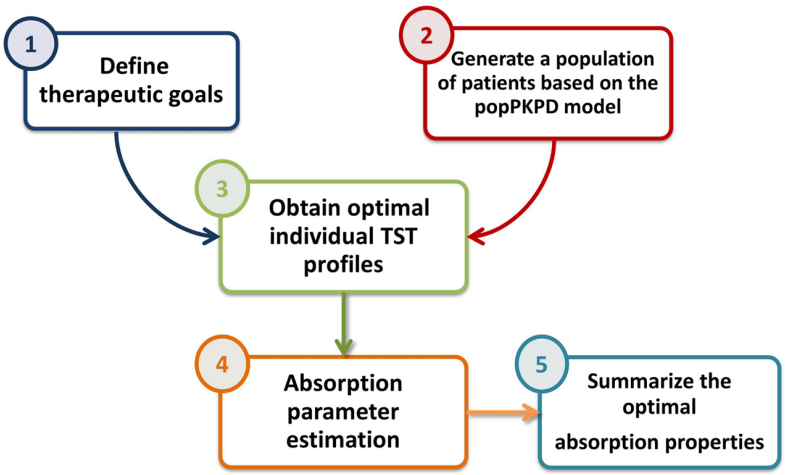


Fig. 2.3 Principal steps implemented in our methodology.

Therefore any optimal control problem can be formulated to find the magnitude of $u(t)$ over the time of study [from initial time t_0 to final time t_f] such that:

$$\begin{aligned} \min_{u(t)} \quad & J(u) = \phi[x(t_f)] + \int_{t_0}^{t_f} L[x(t),u(t)]dt \\ \text{subject to} \quad & \frac{dx(t)}{dt} = f(x(t),u(t),t) \\ & x(t_0) = x_0, \\ & h(x(t),u(t)) = 0, \\ & g(x(t),u(t)) \leq 0, \end{aligned} \tag{2.1}$$

where $J(u)$ is the cost function, $u(t)$ is the control variable; $x(t)$ the vector of state variables; x_0 the set of initial conditions of the state variables; $h()$ the equality constraints; and $g()$ the inequality constraints. The general form of the equation in $J(u)$ is known as Bolza optimization problem [7], which is represented as the sum of a terminal cost functional (Mayer problem) and an integral function of the state and control from t_0 to t_f (Lagrange problem). For a more detailed information see Supplementary Information.

Fig 2.2 shows a schematic representation of the state variables and control input defined in this work. The state system is characterized by the variables that predict serum concentrations of triptorelin (C_{TRP}), concentrations of triptorelin in the shallow and deep peripheral compartments (C_1 and C_2 respectively), drug input profile (D), amount of total receptors (R_T), and optimal testosterone levels (TST), each of them represented by the corresponding ordinary differential equation as shown in Table 2.1. Note that in Fig 2.2 the terms resembling the 0^{th} and 1^{st} order absorption processes have been removed from the original model structure from [3] and have been replaced by the new control

variable $u(t)$. Therefore the expression associated to the rate of change of the levels of TRP in plasma ($\dot{C}_{TRP(t)}$) is:

$$\begin{aligned} \dot{C_{TRP}} = u(t) + \frac{CL_{D1}}{V_{T1}} \cdot c_1(t) + \frac{CL_{D2}}{V_{T2}} \cdot c_2(t) - \frac{CL_{D1}}{V_c} \cdot C_{TRP}(t) \\ - \frac{CL_{D2}}{V_c} \cdot C_{TRP}(t) - \frac{CL}{V_c} \cdot C_{TRP}(t) \end{aligned} \tag{2.2}$$

An additional compartment D was defined, where the dose of TRP administered to the patients (10mg in this evaluation exercise) was placed as initial condition (D_0). The control variable $u(t)$ leaves this compartment and enters to the central (systemic) compartment of TRP as follows:

$$\dot{D} = -u(t) \tag{2.3}$$

Recall that, $u(t)$ (ng/day) does not represent any particular mechanism of absorption (i.e., zero and/or first order kinetics), but a vector of different values that influence the system to behave in a pre-determined (optimal) way.
The aim of this work was to find the time profile of $u(t)$ (input function of TRP into the central compartment) that minimizes an objective (or cost) function and satisfies all constraints which represent the boundaries and therapeutic goals to be achieved (see Table 2.1). The choice of an objective function represents a critical aspect in optimal control problems [8]. Here, the problem is divided into two phases each represented by a different cost function and defined between: (i) 0 and t_{cast} , and (ii) t_{cast} and $\geq 280 + t_{cast}$ days, respectively.

During the first phase (from 0 to t_{cast}) the $u(t)$ profile is optimized to transfer the system from an initial state TST_0(baseline testosterone level) to the final state of 0.5ng/ml (CT value) in the shortest possible time. To solve the first phase of the optimization problem, the following objective function and equality constraint were defined respectively:

$$J_I(u) = t_{cast} \tag{2.4}$$

$$TST(t = t_{cast}) - 0.5 = 0 \tag{2.5}$$

where t_{cast} is an static control variable for minimizing J_I. Here, we wished to obtain the minimum value of t_{cast} that causes TST levels to achieve the CT value. The final time t_{cast} was not known in advance, and that is the reason why the optimization problem was divided into two different phases. Additionally, an inequality constraint was added to limit the initial flare-up of the testosterone below 50% increase with respect to baseline:

$$TST(t) < 1.5 \cdot TST_0 \tag{2.6}$$

Table 2.1 Summary of the setup of the different components of the optimal control problem.

	Phase I, $t \in [0, \mathrm{t_{cast}}]$	Phase II,$t \in [\mathrm{t_{cast}}, 280+\mathrm{t_{cast}}]$
Cost function	$J_I(u) = t_{cast}$	$J_{II}(u) = \int_{t_{cast}}^{280+t_{cast}} TST(t)^2 dt$
State variables		
	$\dot{\mathrm{D}} = -\mathrm{u(t)}$ $\dot{\mathrm{C}}_1 = -\frac{\mathrm{CL_{D1}}}{\mathrm{V_{T1}}} \cdot \mathrm{c_1(t)} + \frac{\mathrm{CL_{D1}}}{\mathrm{V_c}} \cdot \mathrm{C_{TRP}(t)}$ $\dot{\mathrm{C}}_2 = -\frac{\mathrm{CL_{D2}}}{\mathrm{V_{T2}}} \cdot \mathrm{c_2(t)} + \frac{\mathrm{CL_{D2}}}{\mathrm{V_c}} \cdot \mathrm{C_{TRP}(t)}$ $\dot{\mathrm{C}}_{\mathrm{TRP}} = \mathrm{u(t)} + \frac{\mathrm{CL_{D1}}}{\mathrm{V_{T1}}} \cdot \mathrm{c_1(t)} + \frac{\mathrm{CL_{D2}}}{\mathrm{V_{T2}}} \cdot \mathrm{c_2(t)} - \frac{\mathrm{CL_{D1}}}{\mathrm{V_c}} \cdot \mathrm{C_{TRP}(t)} - \frac{\mathrm{CL_{D2}}}{\mathrm{V_c}} \cdot \mathrm{C_{TRP}(t)} - \frac{\mathrm{CL}}{\mathrm{V_c}} \cdot \mathrm{C_{TRP}(t)}$ $\dot{\mathrm{R}}_{\mathrm{T}} = \mathrm{k_{S\_R}(D_R \cdot FDB)} - \mathrm{k_{D\_R}} \cdot \mathrm{R_T(t)}$ $\dot{\mathrm{TST}} = \mathrm{k_{S\_T}}(\frac{\mathrm{AGN+BGN}}{\mathrm{1+AGN+BGN}} \cdot \mathrm{R_T(t)}) + \mathrm{k_{in}} - \mathrm{k_{D\_T}} \cdot \mathrm{TST(t)}$	
Initial conditions		
	$\mathrm{D_I(t_0) = Dose}$ $\mathrm{C_{1_I}(t_0) = 0}$ $\mathrm{C_{2_I}(t_0) = 0}$ $\mathrm{C_{TRP_I}(t_0) = 0}$ $\mathrm{R_{T_I}(t_0) = 1}$ $\mathrm{TST_I(t_0) = TST_0}$	$\mathrm{D_{II}(t_0) = D_I(t_{cast})}$ $\mathrm{C_{1_{II}}(t_0) = C_{1_I}(t_{cast})}$ $\mathrm{C_{2_{II}}(t_0) = C_{2_I}(t_{cast})}$ $\mathrm{C_{TRP_{II}}(t_0) = C_{TRP_I}(t_{cast})}$ $\mathrm{R_{T_{II}}(t_0) = R_{T_I}(t_{cast})}$ $\mathrm{TST_{II}(t_0) = TST_I(t_{cast})}$
Inequality constraints (Boundaries)		
	$\mathrm{D} \in [0, \mathrm{Dose}]$ $\mathrm{TST_I} \in [0.5, 1.5 \cdot \mathrm{TST_0}]$	$\mathrm{TST_{II}} \in [0, 0.5]$
Equality constraints		
	$\mathrm{TST_I(t = t_{cast}) - 0.5 = 0}$	

where $\mathrm{BGN} = \frac{\mathrm{C_{TRP}}}{\mathrm{K_D}}$, $\mathrm{FDB} = \frac{\mathrm{R_T}}{\mathrm{R_{T\_0}}} \cdot (2 - \frac{\mathrm{R_T}}{\mathrm{R_{T0}}})$, $\mathrm{R_{T\_0}} = 1$, $\mathrm{D_R} = \frac{\mathrm{D_{R\_50}}}{\mathrm{D_{R\_50}} + (\frac{\mathrm{AGN+BGN}}{\mathrm{1+AGN+BGN}} - \frac{\mathrm{AGN}}{\mathrm{1+AGN}})}$, $\mathrm{Dose} = 10\mathrm{mg}$ and t_0 makes reference to the initial time. The rest of the parameters have been already defined in Fig 2.2. The subscript $_\mathrm{I}$ and $_\mathrm{II}$ in the model parameters refers to the first and second phase respectively.

The second phase, covering the period between t_{cast} and 280+t_{cast} days, aims to maintain the TST levels below CT. If a second objective function or constraints were not incorporated into the optimization problem, values of TST rose above CT at times much earlier than 280 days. The approach used to overcome the above mentioned undesired effect and maintain TST predictions within the therapeutic goal led to the minimization of a second objective function of the form:

$$J_{II}(u) = \int_{t_{cast}}^{280+t_{cast}} TST(t)^2 dt \tag{2.7}$$

The rationale for formulating J_{II} using a quadratic function ($TST(t)^2$) for minimizing testosterone levels, instead of $TST(t)$, was because it offers relevant mathematical advantages in the context of optimization problems. In optimal control theory, one of the main necessary conditions for optimality is that control variables minimize a Hamiltonian function over $u(t)$. The Hamiltonian becomes convex if quadratic forms are used for the objectives and thus the problem will have a unique minimizer [4]. See Supplementary Information for more information about the Hamiltonian matrix and the necessary and sufficient conditions for optimal control problems. Furthermore, using squared terms amplify the effects of large variations and de-emphasize the contributions of small fluctuations.
Continuity between the two phases of the optimization problem was ensured by imposing the initial conditions of the state variables at phase II to be equal to their final values at the end of the phase I (see Table 2.1).

Alternatively, other objective functions or constraints could have been defined. For example, an alternative approach to model the second phase of the optimal control problem is to only add the inequality constraint $\mathrm{TST}[t_{cast} : (280 + t_{cast})] - 0.5 < 0$, instead of a second objective function J_{II}. This approach resulted in TST levels closer to CT compared to the values obtained with the addition of J_{II}. However, in the work from [9, 1] suggested that a CT value lower than 0.2 ng/ml could be an even better target to maximize therapeutic outcomes of prostate cancer patients. Due to these variations in the definition of the most appropriate CT value, we prioritized the minimization of TST levels during the second phase using J_{II} because we obtained the lowest possible values of TST.

Table 2.1 summarizes the setup of the different components of the optimal control problem described above. There exists different methods to solve this type of problems [10, 11]. In our case, the dynamic optimization problem was solved numerically via direct methods with the IPOPT Solver (Interior Point OPTimizer) [12] which is freely available in the APMonitor Optimization Suite (http://apmonitor.com/) through MATLAB programming environment [13]. The results were evaluated by calculating the proportion of individuals that achieved the described therapeutic goals and constraints.

For more information about control theory see the works from [14, 6, 8] and for a more comprehensive overview of the role of optimal control in cancer research read the reviews from [15, 10, 4].

2.2.3 Mechanistic characterization of the optimal absorption profiles

During the optimal control exercise, values of TST in plasma were obtained approximately every 12h for the first phase and every 120h in the second phase. Given the fact that the disposition, pharmacodynamics, and system parameters were already known as they were randomly generated as described in 2.2.1, the analyses of the TST profiles described in this section focused on the mechanistic/parametric characterization of the absorption process of TRP aiming to provide biopharmaceutics with metrics useful to guide the development of new sustained release formulations. Those metrics are the fractions of the total dose injected absorbed following 0^{th} and 1^{st} order processes, the cumulative drug release profiles over time, the percentage of the dose that should remain in the site of injection at t_{cast} and the time at which the different absorption mechanism are activated.

The absorption model used to estimate the corresponding absorption parameters allowing afterwards computation of metrics is represented in the work from [3] and comprises three non-simultaneous absorption mechanisms, two of them following 1^{st} order kinetics and the third one following a 0^{th} order process. This model is considered of a sufficient complexity to deal with almost any absorption profile that can take place after administration of SR formulations [16, 17]. A schematic representation of the structural model with the corresponding ordinary differential equations is provided in supplementary figure S2.1.

The analyses were performed with the NONMEM version 7.2 software [5], following a two stage approach in which the parameters of each subject are first obtained and summary statistics (median, and 95th confidence intervals) are then calculated. BSV in the absorption parameters was modelled exponentially as described in section 2.2.1 for the rest of model parameters. TST concentrations obtained in step 2.2.2 were logarithmically transformed for the analysis, and residual variability was modeled by using an additive error model on log-transformed data.

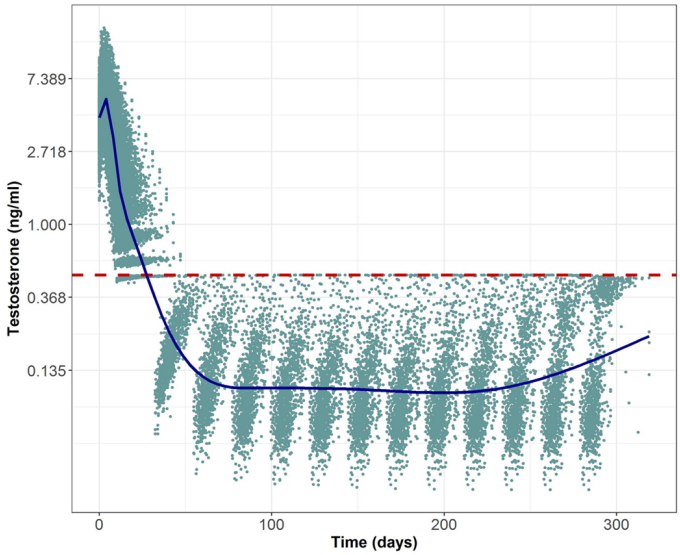


Fig. 2.4 Optimal testosterone (TST) profiles for 1000 simulated individuals. Solid circles represent optimal TST observations obtained after the optimal control approach, solid line represent the median tendency of the data and red dashed line indicates the castration limit (0.5ng/ml) of prostate cancer patients.

2.3 Results

2.3.1 Optimal Pharmacodynamic profiles

Fig 2.4 (blue points) illustrates the optimal testosterone profiles for the 1000 hypothetical individuals that we obtained after applying the optimal control problem formulated in Table 2.1. The initial dose was considered to be 10mg. The code and data to reproduce these results in MATLAB can be found in the online version of the publication. All of them achieved the 3 quantitative therapeutic goals (95% interval confidence between parenthesis) defined in the Introduction section: time to castration was minimized to 18.96 days (11.408 - 36.289), the increase of TST levels at the flare was always smaller than 50% with respect to baseline (36.8%-50.002%), and t_{effect} was greater than 280 for all the patients.

These profiles were generated with the manipulable variable $u(t)$ which could take any values in order to minimize the multi-objective problem. However if we looked to the TRP concentration vs time profiles that induced the optimal TST levels (data not shown), those profiles did not seem attainable by using simple first or zero order kinetics. That was the reason to directly approximate the TST levels with the PKPD model presented by [3] and estimate the most adequate absorption parameters.

2.3.2 Optimal release characteristics

The optimal release characteristics corresponding to the selected PKPD model from [3] are listed in Table 2.2. The final model adequately described the optimal TST profiles calculated in the previous section as shown in the individual profiles of Fig 2.5. A lag time was associated with one absorption

Table 2.2 Population absorption parameters estimated for the optimal triptoreline profiles. The median values and the 95% confidence intervals (CI%) are shown for a population of 1000 patients.

Parameter	Estimate (CI%)
D_{inf} (day)	1.66 (-)
$K_{A1}(day^{-1})$	0.25 (0.108-0.502)
$K_{A2}(day^{-1})$	0.003 (0.0014-0.006)
F_1	0.298 (0.067-0.6)
F_2	0.664 (0.378-0.923)
F_{inf}	0.039 (7.85e-05-0.07)
t_{lag} (day)	3.3 (2.336-5.188)

where D_{inf} is the duration of the zero-order absorption process, K_{A1} and K_{A2} are the first order rate constants of the first and second depot compartments respectively, F_1, F_2 and F_{inf} represent the fraction of the drug associated with the first and second depot compartments and the zero-order absorption process respectively, and t_{lag} is the lag time associated to the first absorption compartment.

compartment. The first order rate constant of absorption of the second depot compartment (K_{A2}) had a very low median value (0.003 day^{-1}), resulting in a slow decay of TRP in serum concentrations. The first order rate constant of the first depot compartment (K_{A1}), instead, had a higher value (0.25 day^{-1}) to allow for a rapid decay of the TST levels in the firsts days of treatment. Table 2.2 also indicates that most of the drug is released following 1^{st} order kinetics as the fraction of drug associated with the 0^{th} order absorption process (F_{inf}) is very small (4%). This result is also reflected in Fig 2.6, where the median tendency of the drug release following each of the absorption mechanisms for the 1000 individuals is shown. The values of the duration of the 0^{th} order process (D_{inf}) varied immensely between individuals (from hours to more than 200 days), thus the variability term was removed from this parameter.

The therapeutic objectives obtained were compared to those from the optimal TST profiles (Fig 2.7) and the 95% confidence intervals were also calculated from the 1000 samples. Minimal time to achieve castration levels was 19.5 days (11.4-56.7) and the median percentage of drug consumed until that moment was 38.9% (16.55%-66.96%). In Fig 2.7A the distribution of t_{cast} values for the 1000 individuals can be appreciated. With the modeling approach 63.9% of the patients had a t_{cast} smaller than 21 days, whereas in the optimal TST profiles this value was equal to 70.7%. Regarding the second therapeutic goal, the initial peak in the TST levels had a median of 55% (17%-75%) increase with respect to baseline. This indicated that the second objective was not always achieved, contrary to the case of the TST profiles obtained by the optimal control problem where the flare up was much more controlled (Fig 2.7B). Nonetheless, the median value was very close to the optimal value of 50%, so we assumed that the modeling approach managed to achieve the second therapeutic goal as well. Finally, the long-term castration had a median value of 351 days (235.9 - 708), which was higher than expected (Fig 2.7C), but we again needed to take into account that a small fraction of

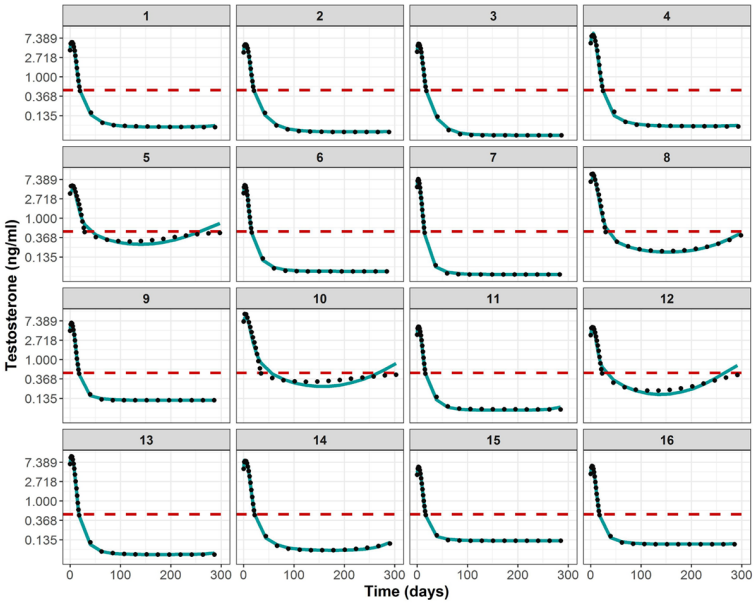


Fig. 2.5 Optimal testosterone profiles of the 1000 virtual patients. Optimal testosterone observations (solid circles) with individual predictions (solid blue lines) of the pharmacokinetic/pharmacodynamic model and a red dashed line indicating the castration limit (0.5ng/ml) of prostate cancer patients.

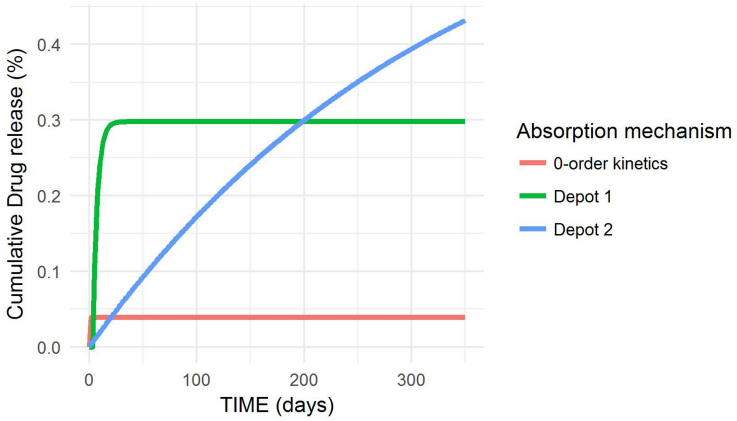


Fig. 2.6 Optimal drug release characteristics following each of the absorption mechanisms.

individuals (7.8%) did not achieve a $t_{cast} + t_{effect}$ above 9 months due to their specific physiological characteristics.

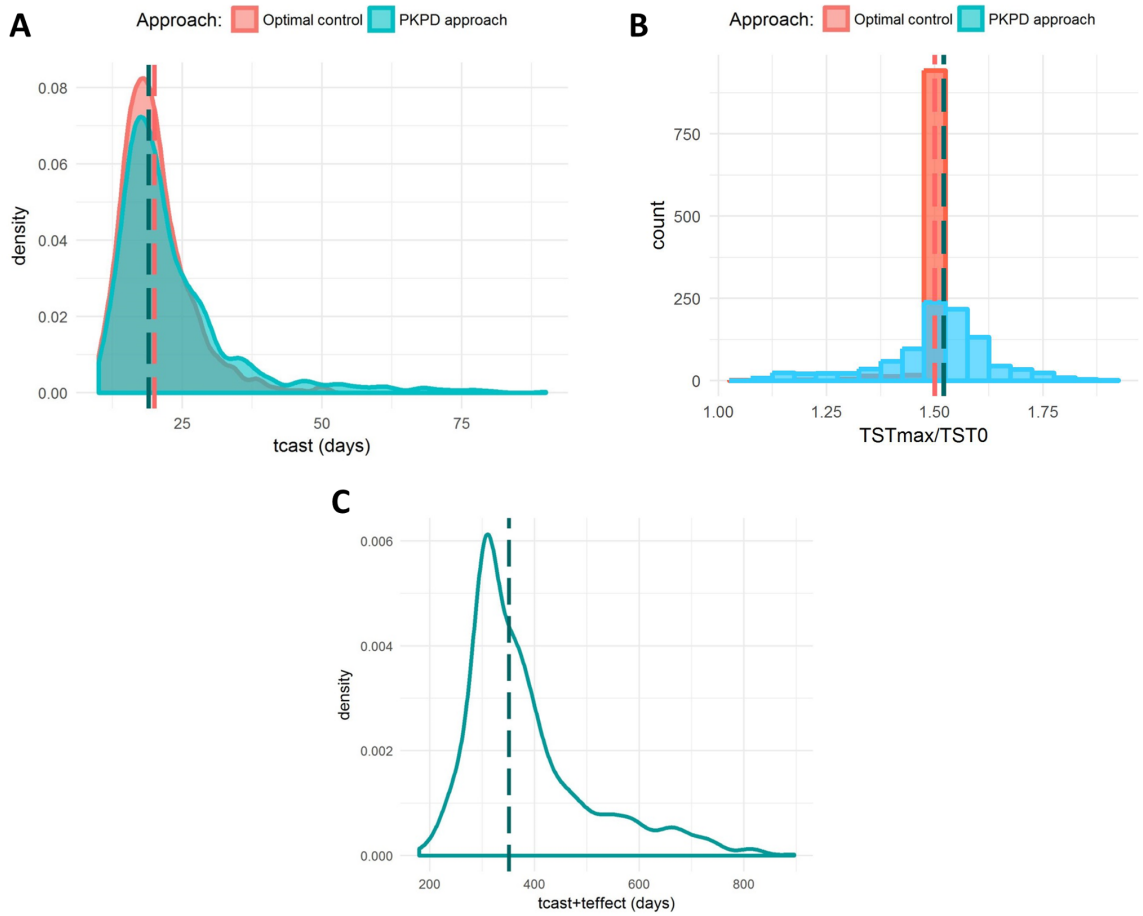


Fig. 2.7 Comparison of the three therapeutic objectives between the optimal control strategy (salmon) and the pharmacokinetic/pharmacodynamic modeling approach (blue) for 1000 individuals. A) Distribution of t_{cast} (time to obtain testosterone levels below castration limit) values. B) Distribution of the values of the testosterone flare-up (maximum testosterone level/baseline testosterone level). C) Distribution of $t_{cast} + t_{effect}$ (castration time after injection of the drug) time values of the modelling approach.

2.4 Discussion

In this paper, we have applied mathematical modeling and control theory to establish optimal drug input (release) profiles to support the development of new release formulation of triptorelin aiming to improve patient coverage. Optimal control has a long and successful history of applications in engineering [8, 18] and economics [19, 20] but also has become an important issue in biomedical research. Especially in clinical cancer research, a significant amount of effort has been devoted to developing mathematical models to identify the most effective chemotherapeutic administration regimens using OC methods [15, 21] and references therein. One of the earliest studies where chemotherapy treatment planning was defined as an OC problem was in the work from [22]. The authors applied OC to improve the treatment administration of the bone cancer IgG multiple myeloma. Understanding the dynamics of resistance mechanisms against chemotherapy and targeted drugs and emerging of adverse effects represent challenges that have also been addressed through these techniques as shown in the works from [23–26]. For example, in [24] the authors added the pharmacokinetics of the drug in the OC problem in order to provide chemotherapeutic protocols in qualitative terms. The injected drug concentration is used as the control variable and the minimization of the number of tumor cells at the end of the treatment is defined as the cost function of the problem. The results showed that the best strategy corresponds to the maximum rate of drug injection when growth rate is assumed to be constant, but not in other type of models.

Combination of different active compounds, are the rule rather than the exception in oncology and other therapeutic areas, but the potentially high number of different possible combinations (including different dosing schemes), makes drug selection an unaffordable task from an experimental trial and error perspective. Therefore any guide on how to administer these therapies to achieve the best possible responses is of great potential as shown by [27, 28].

In the area of infectious diseases, the works from [29] and [30] showed that using treatment regimens obtained from optimal control could lead to a substantial improvement in HIV patients outcome in comparison to the administration of constant-dose standard regimens. Anesthesia is another medical area where optimal control strategies are used to maintain patient response within the desired therapeutic window for the case of the Bispectral Index (BIS) as an indicator of sedation, or the degree of neuromuscular blockade [31, 32].

Despite the approach shown in the current evaluation is not novel in the drug delivery arena, it has been seldom used beyond optimizing drug exposure. Especially in the context of multi-objective optimization, the current example and others presented below indicate that the optimal control approach should be considered part of the computational modeling arsenal advocated by the FDA promoted critical pathway initiative under the Model Informed Drug Discovery and Development (MID3)

paradigm [33].

Here we focused on the non-trivial problem of simultaneously achieving multiple therapeutic goals related to drug onset and offset in the context of clinical trials with a minimum duration of 9 months, which implies high cost and uncertainty regarding the final response outcome. Therefore the possibility of providing to pharmaceutical technology scientists guidance in the form of release/input (absorption) profiles represents a real added value to avoid failed clinical studies. In this context we performed a reverse engineering exercise interpreting the empirical input profiles from a mechanistic biopharmaceutic perspective and showing the practical application of our optimal control analysis, encouraging cooperation between computational and experimental/technology scientists.

In the current exercise the majority of the system accounting for the relationship between dose and response (drug disposition, receptor interaction, and down-regulation mechanisms) was already well characterized with the corresponding typical values and associated variability reported in [3]. In addition the subcutaneous route of administration represents a much simpler biological system compared for example with the oral route. Therefore the optimal control approach used here represents an appropriate choice even recognizing that in more complex situations the advantages offered by similar approaches like non-linear model predictive control [34, 13] could represent a better alternative.

In this context, the main critical aspect of the analysis is the choice of the appropriate structure of the cost function to be minimized and the constraints of the problem. As highlighted in Materials and methods section, we divided the problem into two phases because the minimum time to achieve CT values (t_{cast}) in TST levels was not known in advance. For the first phase, we implemented the Mayer form of optimization problems whereas for the second phase a Lagrange term was used (see Table 2.1). In this work, we focused on the resulting testosterone levels of the prostate cancer patients, but, as shown in the above paragraphs of the Discussion, for other therapeutic areas different objectives could have been established, like the minimization of the tumor cell population at the end of treatment, the maximization of the number of healthy immune system cells or the penalization of excessive application of therapeutic agents [35–37].

The concept of optimization is present at every stage of the drug development process. Optimal design methods, based on the D-optimality criteria which relies on the maximization of the determinant of the Fisher information matrix [38, 39] is becoming also popular to select the appropriate number of subjects in each cohort of the trials, the sampling times and the number of dose levels [40]. However, we must not confuse optimal design methods with optimal control techniques. The aim of the first is to simplify population trials but maintaining the same efficiency as the original studies. Here, they do not alter the system equations nor the objective function of the algorithms and the focus is to search for similar results to the original study (identical pk/pk parameters, similar concentration vs time profiles...). For example, in the work from [40], they used this method to optimize a population

pharmacodynamic experiment of the effect of ivabradine on exercise-induced tachycardia. On the other side, in optimal control, we introduce what are known as control variables into the model equations in order to manipulate the system response towards the desired goal. Therefore, in this type of problems, system equations can be modified and objective functions and constraints defined to search for improved solutions compared to the ones of the original study. Still, both approaches have something in common; they avoid the use of intensive computer simulations when the optimal solutions of a problem are being explored.

2.4.1 Conclusion

Optimal control theory has been applied to a population pharmacokinetic/pharmacodynamic model to derive the optimal drug release profiles to achieve multiple therapeutic goals. The optimal control analysis is more relevant in physiological systems with complex dynamics where simple simulation tuning parameters exercises are not effective to obtain the optimal profiles. Moreover, the flexibility of the method allows to deal with multiple and tight therapeutic objectives performing real optimization. In this context the question of how to define the objective functions and how to quantify our therapeutic goals becomes crucial. Here, we focused on the resulting testosterone levels of the patients, however, within the oncology area, different therapeutic objectives can be established with the goal of improving drug combinations, help to lessen the side effects of cancer treatments, etc.

Finally, the optimal release characteristics have been described based on standard absorption PK models. Although there are some discrepancies between the resulting TST profiles from the optimal control strategy and the modeling approach (see Results), we note that the important aspect of this work was to find the optimal release characteristics for prostate cancer patients, not to perform an ideal PKPD modeling exercise as there was not real data to fit. We conclude that this objective is achieved and that the information summarized in this article could be very useful for the development of new formulations, since it provides insight into the desired absorption characteristics and could produce a broad benefit for future prostate cancer patients.

References

[1] Oefelein MG, Resnick MI. Effective testosterone suppression for patients with prostate cancer: is there a best castration? Urology. 2003;62(2):207–213.

[2] Scher HI, Halabi S, Tannock I, Morris M, Sternberg CN, Carducci MA, et al. Design and End Points of Clinical Trials for Patients With Progressive Prostate Cancer and Castrate Levels of Testosterone: Recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol. 2008;26(7):1148–1159.

[3] Romero E, Vélez de Mendizabal N, Cendrós JM, Peraire C, Bascompta E, Obach R, et al. Pharmacokinetic/pharmacodynamic model of the testosterone effects of triptorelin administered in sustained release formulations in patients with prostate cancer. J Pharmacol Exp Ther. 2012;342(3):788–798.

[4] Schättler H, Ledzewicz U. Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods. Interdisciplinary Applied Mathematics. Springer New York; 2015.

[5] Bauer R. NONMEM users guide introduction to NONMEM 7.2. 0. ICON Development Solutions Ellicott City, MD. 2011;.

[6] Stengel RF. Optimal Control and Estimation. Courier Corporation; 2012.

[7] Cesari L. Lagrange and Bolza Problems of Optimal Control and Other Problems. In: Optimization—Theory and Applications. Applications of Mathematics. Springer, New York, NY; 1983. p. 196–205.

[8] Parkinson AR, Balling R, Hedengren JD. Optimization Methods for Engineering Design. Brigham Young University. 2013;5.

[9] Oefelein MG, Feng A, Scolieri MJ, Ricchiutti D, Resnick MI. Reassessment of the definition of castrate levels of testosterone: implications for clinical decision making. Urology. 2000;56(6):1021–1024.

[10] Shi J, Alagoz O, Erenay FS, Su Q. A survey of optimization models on cancer chemotherapy treatment planning. Ann Oper Res. 2014;221(1):331–356.

[11] Drag P, Styczeń K, Kwiatkowska M, Szczurek A. A Review on the Direct and Indirect Methods for Solving Optimal Control Problems with Differential-Algebraic Constraints. In: Recent Advances in Computational Optimization. Studies in Computational Intelligence. Springer, Cham; 2016. p. 91–105.

[12] Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program. 2005;106(1):25–57.

[13] Hedengren JD, Shishavan RA, Powell KM, Edgar TF. Nonlinear modeling, estimation and predictive control in APMonitor. Comput Chem Eng. 2014;70:133–148.

[14] Bryson AE. Applied Optimal Control: Optimization, Estimation and Control. CRC Press; 1975.

[15] Swan GW. Role of optimal control theory in cancer chemotherapy. Math Biosci. 1990;101(2):237–284.

[16] Tornøe CW, Agersø H, Senderovitz T, Nielsen HA, Madsen H, Karlsson MO, et al. Population pharmacokinetic/pharmacodynamic (PK/PD) modelling of the hypothalamic–pituitary–gonadal axis following treatment with GnRH analogues. Br J Clin Pharmacol. 2007;63(6):648–664.

[17] Buil-Bruna N, Garrido MJ, Dehez M, Manon A, Nguyen TXQ, Gomez-Panzani EL, et al. Population Pharmacokinetic Analysis of Lanreotide Autogel/Depot in the Treatment of Neuroendocrine Tumors: Pooled Analysis of Four Clinical Trials. Clin Pharmacokinet. 2016;55(4):461–473.

[18] Longuski JM, Guzmán JJ, Prussing JE. Optimal Control with Aerospace Applications; 2014.

[19] Sethi SP, Thompson GL. Optimal Control Theory: Applications to Management Science and Economics. Springer Science & Business Media; 2006.

[20] Black F, Scholes M. The Pricing of Options and Corporate Liabilities. J Polit Econ. 1973;81(3):637–654.

[21] Martin R, Teo KL. Optimal Control of Drug Administration in Cancer Chemotherapy. World Scientific; 1994.

[22] Swan GW, Vincent TL. Optimal control analysis in the chemotherapy of IgG multiple myeloma. Bull Math Biol. 1977;39(3):317–337.

[23] Martin RB, Fisher ME, Minchin RF, Teo KL. Optimal control of tumor size used to maximize survival time when cells are resistant to chemotherapy. Math Biosci. 1992;110(2):201–219.

[24] Costa MI, Boldrini JL, Bassanezi RC. Drug kinetics and drug resistance in optimal chemotherapy. Math Biosci. 1995;125(2):191–209.

[25] Coldman AJ, Murray JM. Optimal control for a stochastic model of cancer chemotherapy. Math Biosci. 2000;168(2):187–200.

[26] Kimmel M, Swierniak A. Control Theory Approach to Cancer Chemotherapy: Benefiting from Phase Dependence and Overcoming Drug Resistance. In: Tutorials in Mathematical Biosciences III. Lecture Notes in Mathematics. Springer, Berlin, Heidelberg; 2006. p. 185–221.

[27] de Pillis LG, Fister KR, Gu W, Head T, Maples K, Neal T, et al. Optimal Control of Mixed Immunotherapy and Chemotherapy of Tumors. J Biol Syst. 2008;16(01):51–80.

[28] Ledzewicz U, Moore H. Dynamical Systems Properties of a Mathematical Model for the Treatment of CML. NATO Adv Sci Inst Ser E Appl Sci. 2016;6(10):291.

[29] Gu W, Moore H. Optimal therapy regimens for treatment-resistant mutations of HIV; 2006.

[30] Kirschner D, Lenhart S, Serbin S. Optimal control of the chemotherapy of HIV. J Math Biol. 1997;35(7):775–792.

[31] Almeida J, Mendonça T, Rocha P. A simplified control scheme for the Depth of Anesthesia. IFAC-PapersOnLine. 2016;49(5):230–235.

[32] Almeida J, Mendonça T, Rocha P. A Simplified Control Approach for the Neuromuscular Blockade Level. In: CONTROLO 2016. Lecture Notes in Electrical Engineering. Springer, Cham; 2017. p. 37–44.

[33] Woodcock J, Woosley R. The FDA Critical Path Initiative and Its Influence on New Drug Development. Annu Rev Med. 2008;59(1):1–12.

[34] Grüne L, Pannek J. Nonlinear Model Predictive Control. In: Nonlinear Model Predictive Control. Communications and Control Engineering. Springer, London; 2011. p. 43–66.

[35] Stengel RF, Ghigliazza RM, Kulkarni NV. Optimal enhancement of immune response. Bioinformatics. 2002;18(9):1227–1235.

[36] Engelhart M, Lebiedz D, Sager S. Optimal control for selected cancer chemotherapy ODE models: a view on the potential of optimal schedules and choice of objective function. Math Biosci. 2011;229(1):123–134.

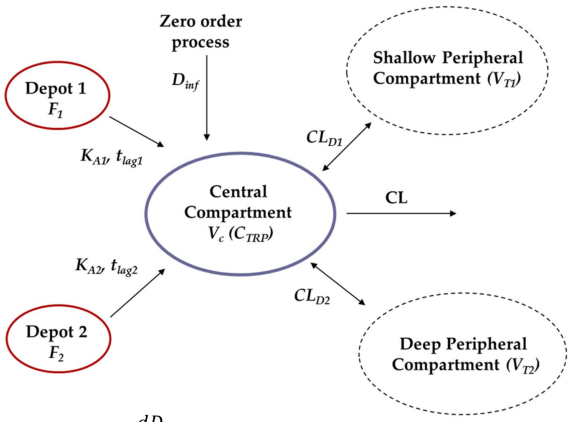
[37] He Q, Zhu J, Dingli D, Foo J, Leder KZ. Optimized Treatment Schedules for Chronic Myeloid Leukemia. PLoS Comput Biol. 2016;12(10):e1005129.

[38] Fedorov VV. Theory of optimal experiments. Elsevier; 1972.

[39] Mentré F, Mallet A, Baccar D. Optimal design in random-effects regression models. Biometrika. 1997;84(2):429–442.

[40] Duffull SB, Mentré F, Aarons L. Optimal design of a population pharmacodynamic experiment for ivabradine. Pharm Res. 2001;18(1):83–89.

Supplementary Material



$$\frac{dD_2}{dt} = -K_{A2} \cdot D_2 \cdot F_2 \frac{dD_1}{dt} = -K_{A1} \cdot D_1 \cdot F_1$$

$$\frac{dC_{TRP}}{dt} = \left(K_{A1} \cdot D_1 \cdot F_1 + K_{A2} \cdot D_2 \cdot F_2 + \left(\frac{Dose}{D_{inf}}\right) \cdot F_{inf} + \left(\frac{CL_{D1}}{V_{T1}}\right) \cdot c1 - \left(\frac{CL_{D1}}{Vc}\right) \cdot Ctrp + \left(\frac{CL_{D2}}{V_{T2}}\right) \cdot c2 - \left(\frac{CL_{D2}}{Vc}\right) \cdot C_{TRP} - \left(\frac{CL}{Vc}\right) \cdot C_{TRP}\right)$$

$$\frac{dC_1}{dt} = -\left(\frac{CL_{D1}}{V_{T1}}\right) \cdot c1 + \left(\frac{CL_{D1}}{Vc}\right) \cdot C_{TRP}$$

$$\frac{dC_2}{dt} = -\left(\frac{CL_{D2}}{V_{T2}}\right) \cdot c2 + \left(\frac{CL_{D2}}{Vc}\right) \cdot C_{TRP}$$

$$F_{inf} = 1 - F_1 - F_2$$

Fig. S2.1 Original pharmacokinetic model for triptorelin from [3]. D_{inf} is the duration of the zero-order absorption process; K_{A1} and K_{A2} are the first order rate constants of the first and second depot compartments respectively; F_1, F_2 and F_{inf} represent the fraction of the drug associated with the first and second depot compartments and the zero-order absorption process respectively; t_{lag1}/t_{lag2} is the lag time associated to the first/second absorption compartment; C_{TRP} is the serum concentrations of Triptorelin; CL, the apparent total clearance; V_c, V_{T1}, and V_{T2}, apparent volumes of distribution of the central, shallow, and deep peripheral compartments respectively; and CL_{D1} and CL_{D2}, distribution clearances between the central and peripheral compartments.

S1 Data: APMonitor MATLAB code and datasets. APMonitor Optimization Suite code for MATLAB and datasets to reproduce the results of this work can be found in the online version of the publication.

Supporting information

Here, we wanted to summarize the mathematics involving the process of finding optimal control histories for dynamic systems. These methods are analogous to the static optimizations, but the computations are necessarily more complex. Additionally, we consider the method of Pontryagin maximum principle to address optimal control problems having path constraints.

DYNAMIC SYSTEMS

In dynamic systems, state equations that describe the system and controls are expected to evolve over time and not necessarily remain at a steady state.
Consider the dynamic system described by the ordinary differential equation of the form:

$$\dot{x}(t) = f(x(t), u(t)), \mathrm{x}(t_0) = x_0 \tag{i}$$

given the control variable $u(t)$ for $t_0 \leq t \leq t_f$ and the following cost function over the time interval [t_0 , t_f]:

$$J = \varphi\left[x\left(t_f\right)\right] + \int_{t_0}^{tf} L\left[x\left(t\right), u\left(t\right)\right]\, dt \tag{ii}$$

The first term of the cost function ii is known as Mayer problem and is a scalar algebraic function of the final state and time and works as a terminal penalty. It is used to assure that one or more state variables satisfy some condition at the end of the trajectory. The second term is known as Lagrange problem and is an integral function of the state and control from t_0 to t_f. When both terms are considered in the cost function, the problem is known as Bolza and its general form is shown in equation ii (Stengel 1986; Betts 2001).
In optimal control problems, the control history $u(t)$ that minimizes J is to be found, subject to specific constraints.

DYNAMIC OPTIMIZATION WITH DYNAMIC CONSTRAINTS

As previously described, generally dynamic constraints take the form of ordinary differential equations. To guarantee that the dynamic constraint is considered in the minimization of a cost function, is has to be adjoined to the integrand term of the equation in such a way that the numerical value of the cost function remains unchanged at its optimum.
Note that in **static minimization**, an equality constraint of the form f(x,u)=0 can be adjoined to the cost function J using Lagrange multipliers (vector λ) to form the augmented cost function J_A:

$$J_A = J + \lambda^T f(x,u) \tag{iii}$$

J_A is identical to J at the minimum because $f(x,u)$ must be zero at that point.

In dynamic optimization however, the equality constraint is derived from the system's differential equation $\dot{x}(t) = f(x(t), u(t))$, which is equivalent to $f(x(t), u(t)) - \dot{x}(t) = 0$. If we convert the constrained optimal control problem into an unconstrained optimal control problem using the Lagrange multipliers, the augmented cost function J_A has the following form:

$$J_A = \varphi[x(t_f)] + \int_{t0}^{tf} \{L[x(t), u(t)] + \lambda^T(t)[f[x(t), u(t)] - \dot{x}(t)]\}\, dt \tag{iv}$$

This motivates the construction of the Hamiltonian function defined as

$$H[x(t), u(t), \lambda(t)] = L[x(t), u(t)] + \lambda^T(t) f[x(t), u(t)] \tag{v}$$

Thus,

$$J = \varphi[x(t_f)] + \int_{t_0}^{tf} \left\{H[x(t), u(t), \lambda(t)] - \lambda^T(t)\, \dot{x}(t)\right\}\, dt \tag{vi}$$

Now J_A depends on $\dot{x}(t)$. Using integration by parts,

$$\int_{t_0}^{t_f} \lambda^T(t)\, \dot{x}(t) = \lambda^T(t_f) x(t_f) - \lambda^T(t_0) x(t_0) - \int_{t_0}^{t_f} \dot{\lambda}^T(t) x(t) \tag{vii}$$

the cost function becomes

$$J_A = \varphi[x(t_f)] + \left[\lambda^T(t_0) x(t_0) - \lambda^T(t_f) x(t_f)\right] + \int_{t_0}^{tf} \left\{H[x(t), u(t), \lambda(t)] + \dot{\lambda}^T(t)\, x(t)\right\}\, dt \tag{viii}$$

From this point, the augmented cost function will be written simply as J.

CONDITIONS FOR OPTIMALITY

One of the main issues in optimal control is to assure whether an optimal control exists for a given problem. In the next section, we describe a set of conditions which any optimal control must satisfy.

Necessary conditions for optimality

Necessary conditions indicate that the cost function should be insensitive to small control variations (Δu) in the optimum point. This is mathematically expressed as $\Delta J^*[u*, \Delta u] = 0$ (the first variation of J due to small variation in u must be equal 0). Setting the terms that multiply these variations to be zero yields to the three necessary conditions for optimality, known as the Euler-Lagrange equations (described in detail in (Stengel 1986)):

$$1. \left[\frac{\partial \varphi}{\partial x} - \lambda^T\right]_{t=t_f} = 0 \tag{ix}$$

$$2.\left[\frac{\partial H}{\partial x}+\dot{\lambda}^{T}\right]=0 \qquad in\ (t_0,t_f) \tag{x}$$

$$3.\frac{\partial H}{\partial u}=0\ \ in\ (t_0,t_f) \tag{xi}$$

Note that, in this section, we consider optimal control problems having no restriction on the control variables.

Sufficient conditions for optimality

In many optimal control problems, cost stationarity may imply optimality because there is sufficient knowledge of the system and flexibility in the choice of the cost function. Even so, in (Stengel 1986) three sufficient conditions are described (convexity condition, normality condition and uniqueness or Jacobi condition), which along with the necessary conditions, would assure optimality. The most important one (and the only condition that is discussed for this work) is the convexity condition, also known as Legendre-Clebsch condition:

$$\left[\frac{\partial^2 H(x^*,\ u^*,\ \lambda^*)}{\partial u^2}\right]>0 \quad in\ (t_0,t_f) \tag{xii}$$

Where x*, u* and λ* refer to the optimal state, control and Lagrange variables found as solution to the problem respectively. This property implies that every stationary point in u is a strict local minimizer of the Hamiltonian H in the control. Using quadratic formulations in the cost function considerably simplifies the analysis of this condition because the Hamiltonian becomes convex in the control u and thus has a unique minimizer. While this does not guarantee that the controls found by the analysis are necessarily optimal, it introduces important mathematical advantages (Schättler and Ledzewicz 2015).

Pontryagin's Principle

In this section, we present more general necessary conditions for optimality for those problems having bounded control magnitudes. Such conditions were generalized by Pontryagin and co-workers in their Maximum principle (sometimes also referred as Pontryagin's Minimum Principle) (Pontryagin et al. 1962).

Because of the restriction on u, the necessary condition of section 3.1 changes to $\Delta J^*[u^*,\Delta u]\geq 0$. This implies that

$$H\left[x^*(t),u^*(t),\lambda^*(t)\right]\leq H\left[x^*(t),u^*(t)+\Delta u,\lambda^*(t)\right] \tag{xiii}$$

Equation xiii is the necessary conditions stated by Pontryagin in his theorem, which along with equations ix and x, state the necessary conditions for optimality when control variables are subject to constraints on their magnitudes. The maximum principle can in some cases provide a solution to

problems where the Euler-Lagrange equations fail. One example is when linear terms are used in the cost function and the equation $\frac{dH}{du}$ becomes identically zero.

Both Euler-Lagrange equations and the maximum principle are difficult to apply to large non-linear systems with state or control variable constraints and this may lead to computationally intractable problems. In these cases, where an analytical solution is not possible to be found, computational algorithms are inevitable in solving optimal control problems (Hedengren et al. 2014).

References:

1. Betts, John T. 2001. Practical Methods for Optimal Control Using Nonlinear Programming. Society for Industrial and Applied Mathematics.

2. Hedengren, John D., Reza Asgharzadeh Shishavan, Kody M. Powell, and Thomas F. Edgar. 2014. "Nonlinear Modeling, Estimation and Predictive Control in APMonitor." Computers & Chemical Engineering 70 (November): 133–48.

3. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. Mishchenko. 1962. "The Mathematical Theory of Optimal Processes (International Series of Monographs in Pure and Applied Mathematics)." Interscience, New York.

4. Schättler, Heinz, and Urszula Ledzewicz. 2015. Optimal Control for Mathematical Models of Cancer Therapies: An Application of Geometric Methods. Interdisciplinary Applied Mathematics 42. Springer New York.

5. Stengel, Robert F. 1986. Optimal Control and Estimation. Courier Corporation.

3

A Cancer Evolution Simulation Optimizer

A Cancer Evolution Simulation Optimizer

Itziar Irurzun-Arana[1,2,3], Thomas O. McDonald[3,4,5,6], Iñaki F. Trocóniz [1,2], Franziska Michor[3,4,5,6]

1 Pharmacometrics & Systems Pharmacology group, Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Navarra, Spain.

2 IdiSNA, Navarra Institute for Health Research, Pamplona, Navarra, Spain.

3 Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA.

4 Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA.

5 Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA.

6 Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA.

Manuscript in preparation.

3.1 Introduction

A major obstacle in cancer research is the development of resistance to anticancer drugs. Although this phenomenon has long been studied, it has gained even more attention after the introduction of targeted therapies and technological advances such as next generation DNA/RNA sequencing. Targeted therapy differs from cytotoxic chemotherapy in that it not only leads to more specific effects with reduced toxicity, but also promises a future of personalized tailored treatment. Unfortunately, the initial clinical response to targeted therapies is almost always temporary, as acquired resistance to these drugs invariably develops [25]. Emergence of resistance is also a therapeutic relevant consequence of administering conventional chemotherapeutic agents [57].

Clonal evolution represents a key element in understanding these resistance mechanisms [24]. In this context, cancer cells evolve specific genetic and/or epigenetic alterations during tumor growth that lead to the progression of different sub-clones within a single tumor. The heterogeneity of cancers has important negative implications for targeted therapies because somatic mutations continue to happen even in the presence of drugs and new drug resistant cells with improved fitness and malignant potential may arise. Moreover, drug resistance can also emerge through therapy-induced selection of a small resistant subpopulation of cells that was already present in the original tumor (intrinsic resistance).

Clonal evolution gradually leads to the aggressiveness of the cancer and resistance to treatment. This realization has given rise to many stochastic mathematical models of genetic resistance, i.e. resistance driven by (epi)genetic alterations in cancer cells, since these alterations emerge as random events during cell division [18, 4]. Some of the earliest work about resistance arising due to point mutations was presented in the 1980s by Goldie and Coldman [22, 23, 12], where the authors proposed the use of a class of stochastic models known as branching process models to study pre-existing or acquired resistance to chemotherapy in tumor cell populations.

Branching process models describe the growth and composition of populations by stochastically reproducing individuals [37, 27]. Multi-type branching processes (branching processes where offspring can be of different type than the parent) are convenient for modelling clonal evolution of cancer cells, where new (epi)genetic alterations emerge as random events during cell division and give rise to tumor subclones with different fitness (i.e. different proliferative capacity or resistance to apoptosis) than their ancestors [47]. This approach has inspired other groups to mathematically characterize drug resistance mechanisms and to investigate potential administration schedules that could delay the appearance of this resistance. For instance, Kendall proposed a two-type branching process where wildtype and mutant cell types were assumed to randomly divide, die, and mutate according to different rates [35]. This model is also referred to as a birth-death branching process because it considers that each cell type experiences random birth (cell divides into two offspring and a

mutation is possible to happen) or death. Iwasa and colleagues [32] used a similar two-type birth and death process to model the dynamics of resistance emerging due to genetic alterations in a population of drug sensitive cells prior to the start of therapy. Multi-drug resistance has also been studied using these types of models in the work from Komarova and Wodarz [39, 38]. More recently, Bozic et al. [6] used multi-type branching processes to predict the effects of combination therapies on tumors and showed that drugs given simultaneously have a higher probability of success compared to sequential therapy.

Hence, understanding how resistance to targeted therapy or conventional chemotherapy occurs is necessary to prolong the effect of modern anti-cancer drugs and try to find optimum dosing strategies. Further complicating the issue, the existence of a multitude of resistance mechanisms necessitates administering drugs in combination, significantly complicating the endeavor of identifying optimal intervention strategies. In addition, high doses or long term use of drugs can cause severe adverse effects in the patients and hence the assessment of the dose-limiting toxicities (DLT) and maximally tolerated doses (MTD) is also very relevant when searching for the optimal regimen. A thorough understanding of the important determinants of cancer evolution and possible side effects under single-agent and combination therapies is therefore crucial for correctly predicting treatment outcomes.

We have previously developed a comprehensive computational strategy to explore the evolutionary dynamics of heterogeneous tumor cell populations while taking pharmacokinetic and drug interaction effects into account [16, 17, 7]. This approach is made up of a cell-level description of the changes in sensitive and resistant cells over time and in response to treatment in the form of a birth-death process model of cell growth, death, and mutation acquisition (see quick guide to equations for a general description of this model). It also includes mathematical models to describe the time course of drug concentrations and to characterize the interactions of the drugs. These characteristics represent a departure from previously developed stochastic models of emergence of resistance during therapy which were defined as time-homogeneous processes where the birth and death rates of sensitive and resistant cells were not directly influenced by drug concentrations and were assumed to be constant.

We note that the pharmacokinetic processes included in previous works [17, 15, 41, 7] were simple exponential decay models of drug concentrations. Even in the case where the dose was given orally to the patients, no absorption parameters were estimated because the time scale considered for the dosing strategies (days instead of hours) plus the rapid absorption of the drug allowed to discard this parameter. However, not all the treatments present such a simple concentration-time profile. Additionally, the effect of drug concentrations on the growth kinetics of sensitive and resistant cell populations were considered to be linear [7] despite the fact that *in vitro* growth inhibition response of cell lines to a range of different drug concentrations is generally non-linear and conventionally modeled with a hill function [19, 51].

In this work, we aim to extend this hybrid methodology to account for more complex pharmacokinetic models and non-linear effects of drugs on the growth and death kinetics of cancer cells. Furthermore, we integrate this framework in an R package called ACESO (A Cancer Evolution Simulation Optimizer), providing users with an accessible tool to rationally identify optimum single-agent and combination treatment administration strategies for oncogene-driven cancers. We demonstrate the use of ACESO to explore optimum dosing strategies using publicly available data from the Harvard Medical School (HMS) Library of Integrated Network-based Cellular Signatures (LINCS) Database [34] (http://lincs.hms.harvard.edu/db/). Additionally, we showed how the new features and flexibility of ACESO could improve the predictions of a previous application of this evolutionary framework on the identification of optimum administration schedules for non-small cell lung cancer patients, matching the results obtained in subsequent publications [15, 56]. We conclude that this multi-scale framework represents a crucial step towards making clinically relevant predictions as it considers the most important aspects governing treatment response and cancer evolution.

Quick Guide to Equations and Assumptions

Multi-type branching process

A branching process is a stochastic model of cell division, mutations events and cell death used to describe the growth and composition of tumor cell populations. It is an example of a first order Markov process that models a population in which the probability of observing n cells at time $t+\Delta t$ (Δt being a very short time interval) only depends on the state of the population at time t and not on its earlier history. Multitype branching processes are branching processes where daughter cells can be of different type than the parent (e.g. a genetic alteration occurs that make them more aggressive, metastatic or resistant to a drug).

Mayor assumptions of the model:

- Only 3 different events are possible for each cell in the population: cell division, cell division with mutation or death.
- Each cell is fully described by cell-intrinsic birth, mutation and death rates.
- Cellular events (replication, mutation and death) do not influence each other, so all events occur independently.
- The waiting time between events a cell can undergo is distributed exponentially according to its intrinsic rates.
- Birth and death rates are influenced by the changing drug concentrations over time, which vary in function of the pharmacokinetic model defined, the model parameter values and the dosing time. The mutation rate parameter can also vary over time, but for the equations described below, is assumed to be constant.

Taking into account these assumptions, the expected number of cells at any time after treatment initiation was derived in previous publications (see [7] and its supplementary material) and is given by:

$$\begin{aligned}\langle n_i(t)\rangle &= N_0 \cdot e^{-\int_0^t [b_0(s)\cdot(1-\sum_{z=1}^{N} u_z)-d_0(s)]ds} && for\ i=0\\ &= \frac{N_i + \int_0^t [e^{-\int_0^s [b_i(\tau)-d_i(\tau)]d\tau} \cdot u_i \cdot b_0(s) \cdot N_0 \cdot e^{\int_0^t [b_0(\tau)\cdot(1-\sum_{z=1}^{N} u_z)-d_0(\tau)]d\tau}]ds}{e^{-\int_0^t [b_i(s)-d_i(s)]ds}} && for\ i>0\end{aligned} \tag{3.1}$$

where $i=0$ denotes the original sensitive cell type and $i=1,...,N$ denotes the N different resistant cell types. Birth, death and mutation first order rates are denoted by b_i, d_i and u_i respectively and the initial number of sensitive and resistant cells is represented by N_0 and N_i. Since the first order mutation rate constant is always much less than 1 ($u_i << 1$), the sum

over the mutation rates of all N-resistant types $\sum_{z=1}^{N} u_z$ can be removed from the equation. The expression for the expected number of cancer cells considering cross-resistance to drugs is shown in the Supplementary material.

The probability that there exists at least one resistant cell of any type ($i = 1,...,N$) at time T after treatment initiation was also derived in previous works [17] and is given by:

$$P_R(T) = 1 - e^{\left[-\int_0^T \left(\sum_{i=1}^{N} N_0 \cdot e^{\int_0^t [b_0(\tau)(1-\sum_z u_z) - d_0(\tau)]d\tau} \cdot b_0(t) \cdot u_i \cdot (1 - P_{ext,i}(t,T))\right) dt\right]} \tag{3.2}$$

$$where \quad P_{ext,i}(t,T) \equiv \frac{\int_0^{T-t} d_i(\tau+t) \cdot (e^{\int_0^{\tau} d_i(\eta+t) - b_i(\eta+t) d\eta}) d\tau}{1 + \int_0^{T-t} d_i(\tau+t) \cdot (e^{\int_0^{\tau} d_i(\eta+t) - b_i(\eta+t) d\eta}) d\tau}$$

The expressions above cannot be solved analytically for time-dependent growth and death rates, and therefore has to be solved numerically.

Net growth, birth and death rate estimation from data

In order to estimate the net growth rate of the cells, an exponential growth model based in the following ordinary differential equation (ODE) is defined:

$$\frac{dN}{dt} = \lambda_j \cdot N \tag{3.3}$$

where N is the number of viable cells, λ_j is the first order net growth rate parameter and j represents the different drug concentrations analyzed in the experiment. This model implies that each concentration has its own net growth rate associated. The analytical solution of this equation is:

$$N(t) = N_0 \cdot e^{\lambda_j \cdot t} \tag{3.4}$$

where N_0 is the number of viable cells at time 0.

In order to estimate the death rate from apoptosis assay data where the number of death cells is counted over time and different drug concentrations, the following ODE was defined:

$$\frac{dN_d}{dt} = d_j \cdot N \tag{3.5}$$

where N_d is the total number of death cells and d_j is the first order death rate parameter for the different drug concentrations tested. As the equation for N has been already defined, the

following analytical solution for N_d can be obtained:

$$N_d(t) = \int_0^t (d_j \cdot N_0 \cdot e^{\lambda_j \cdot t})dt = \frac{N_0 \cdot d_j}{\lambda_j} \cdot (e^{\lambda_j \cdot t} - 1) \tag{3.6}$$

As the value of λ_j is already known from the previous exercise and the values for N_0 and N_d over time and different drug concentrations can be obtained from the dataset, the different d_j parameters can be easily obtained using a least square minimization. Finally, as the net growth rate parameter λ_j is the difference between cell proliferation and death, the birth rates b_j for the different drug concentrations can be obtained from the values of λ_j and d_j as $b_j = \lambda_j + d_j$.

Non-linear concentration-response curves

Many drug concentration-effect relationships are described by nonlinear sigmoid models, which are characterized by a sigmoidal or "S" shape. The four-parameter logistic equation (or Hill equation) is one of the most common approaches to describe single-agent concentration-response curves of this type [51]. In this work, the responses being analyzed are the first order birth and death rates of sensitive and resistant cell lines. This model is defined by (the equation is customized for the effect exerted by a drug on the birth rate):

$$B(C) = B_{max} + \frac{B_0 - B_{max}}{1 + (\frac{C}{EC_{50}})^h} \tag{3.7}$$

where B is the first order rate constant obtained at concentration C C, B_{max} is the maximum effect of the drug on the birth rate, B_0 is the intrinsic birth rate when no drug is present, EC_{50} is the inflection point of the curve and represents the concentration corresponding to 50% of the maximum effect and h is the shape parameter linked to the steepness of the curve. The five-parameter logistic model is an extension of this equation for fitting asymmetrical data [51] as it adds an asymmetry factor parameter defined as f in the previous equation:

$$B(C) = B_{max} + \frac{B_0 - B_{max}}{[1 + (\frac{C}{EC_{50}})^h]^f} \tag{3.8}$$

These models can be generalized to any measure of drug exposure (e.g., dose, plasma concentration, or area under the concentration vs time curve).

Apart from these two examples, ACESO incorporates other nonlinear dose-response models like the three and four-parameter Gompertz model, Weibull models, etc. for parameter estimation. A simple linear model can also be fitted to the data.

Compartmental pharmacokinetics/Population pharmacokinetics
Pharmacokinetic (PK) models consist of algebraic or differential equations used to quantitatively describe how drug concentrations change over time in a biological system following its administration. Compartmental PK analysis views the body as consisting of compartments between which drug distributes and from which elimination occurs. The transfer of drug between these compartments is represented by distribution clearances or first order rate constants.

The simplest compartmental model is the one-compartment model. In this model, the body is depicted as a kinetically homogeneous unit, which means that the drug achieves instantaneous distribution throughout the body following administration and that the drug equilibrates instantaneously between tissues. Thus the plasma drug concentration-time profile $C(t)$ shows a simple exponential decay when the drug is given intravenously as a bolus:

$$C(t) = \frac{D}{V} \cdot e^{(-\frac{CL}{V} \cdot t)} \tag{3.9}$$

where D represents the dose of drug given to the patient, V is the apparent volume of distribution and CL the total drug elimination clearance.

The two-compartment model resolves the body into a central compartment and a peripheral compartment between which drug distributes. Here, the drug does not achieve instantaneous distribution following administration into the central compartment and thus the log plasma drug concentration–time profile shows a biphasic response.

These models can be further complicated to allow for more than two compartments, more complicated absorption profiles, delayed responses, etc. An overview of standard pharmacokinetic compartmental models, their ordinary differential equations and concentration-time profiles under single and multiple drug administrations can be found in Figure 5 from the Introduction section of the thesis.

3.2 Material and Methods

3.2.1 General structure of the framework

We model the cancer cell population during treatment with a multi-type time-inhomogeneous branching process (see *Quick Guide to Equations and Assumptions* for a brief description of the model assumptions and mathematical equations). Here we consider only the sub-population of cancer cells that is capable of self-renewal to produce and maintain a resistant cell clone.

The original cells with oncogene-activating mutations (also denoted type 0 cells) are referred to as sensitive cells, since they contain all (epi)genetic alterations necessary for conferring the cancer phenotype, but anti-cancer therapies are effective against them. The sensitive cancer cells proliferate and die with first order rates $b_0(t)$ and $d_0(t)$ respectively, which can be modulated by the cytostatic and/or cytotoxic effect exerted by the drug(s) as a function of their concentration vs time profiles.

Sensitive cells may accumulate different mutations at a first order rate u_i per cell division and generate new clones harboring specific resistance mechanisms. This mutation parameter can be a constant value or be modified (i.e., augmented) as a result of the presence of a drug in the body. The new resistant cell types (denoted type i cells, with $i = 1,2,...$ depending on the number of resistant clones being modeled) are again characterized by their first order birth and death rates $b_i(t)$,$d_i(t)$, which may also show a concentration dependent profile. Each of these resistant cell types can then again mutate to accumulate further alterations and become resistant to more than one drug. For a schematic representation of a two-type branching process with cross-resistance see Figure 3.1A.

This quantitative approach considers both situations in which there is preexisting resistance as well as de novo resistance at the time of tumor diagnosis and treatment. Analytic solutions for the expected number of cells of any type as a function of time and the probability of developing resistance for an arbitrary number of cell types were derived in previous works [17, 7] and are summarized in the quick guide to equation section, allowing for faster predictions without the need for detailed stochastic simulations of the evolutionary process.

This cell-level description of how heterogeneous cell populations evolve over time is then coupled to a pharmacokinetic (PK) model, which describes how the drug concentration changes over time (Figure 3.1B). PK refers to the processes by which drugs are absorbed, distributed and eliminated from the body. In order to implement more complex compartmental PK models and simulate different routes of administration and any type of dosing schedules, we integrated the *mrgsolve* R package [2] in ACESO, which allowed for rapid simulations of ordinary differential equation (ODE) based models due to its implementation in C++ (see *Quick guide to equation and Assumptions* for an introductory

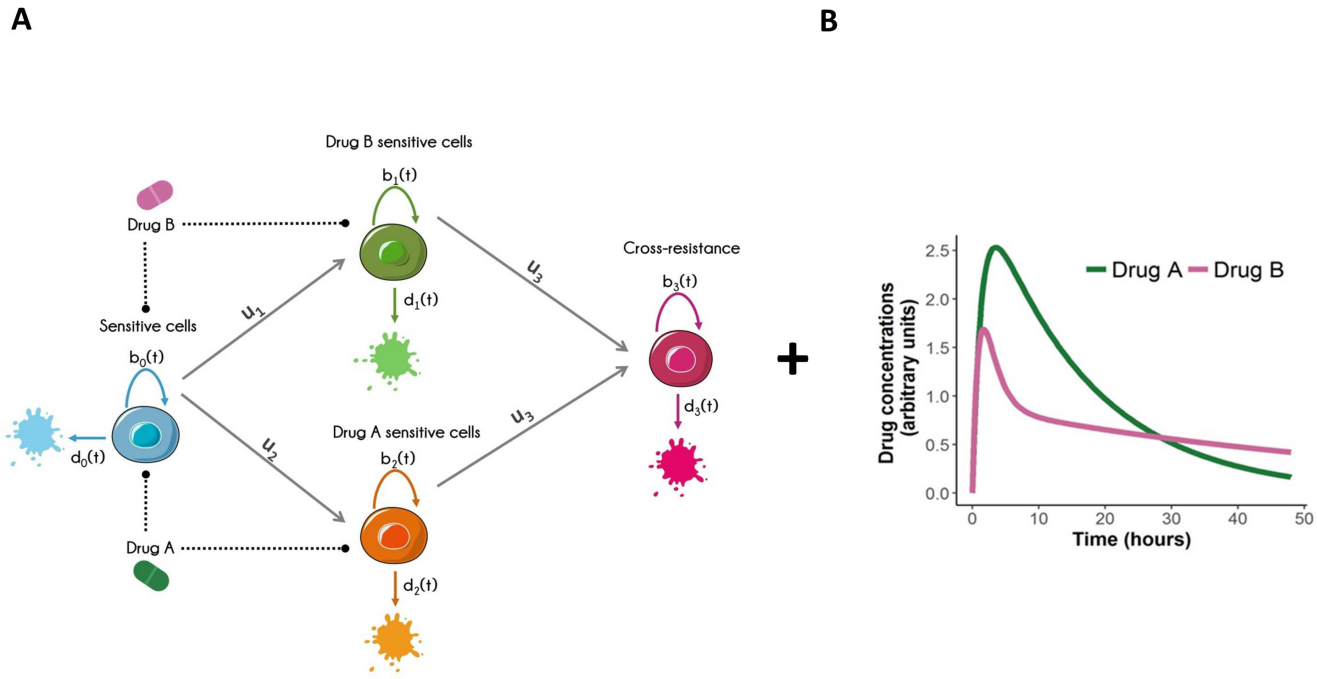


Fig. 3.1 Schematic representation of: A) a two-type branching process with cross-resistance. The type 0 cell (blue), which is sensitive to both drugs A and B, proliferates at rate $b_0(t)$, dies at rate $d_0(t)$ and can accumulate mutations at rate u_i per cell division to give two different cell types, type 1 (green) and 2 (orange), which are sensitive to either drug A or drug B. Type 1 and 2 cells can mutate again and become resistant to both drug A and B (type 3 cells, magenta). Each resistant cell type has its own proliferation and death rates, $b_i(t)$, $d_i(t)$, with $i = 1,2...$ B) Drug concentration-time profile governed by the pharmacokinetic parameters of drug A and B after a single-dose administration.

explanation about compartmental pharmacokinetics).

3.2.2 Model development and outcome

In the following the data analyses and simulations procedures are described. Briefly two types of approximations are combined sequentially as described below. First the parameters driving tumor cell proliferation and death, and pharmacokinetic processes are estimated. Then, those parameters are used to simulate the expected number of different type of cells and probability that there exists at least one resistant cell based on the equations derived from the multitype branching process model shown in the *Quick guide to equations and Assumptions*.

Estimation of birth and death first order rate constants and drug effect parameters

Cell viability and death *in vitro* assays are required for both sensitive and resistance cells lines used in the model to determine the growth and death rate parameters for each cell with and without the presence of drugs. One example of viability assay to characterize cell proliferation is the MTS assay [13]. Here, the number of viable cells is measured over time which is then used to identify a first order

proliferation rate constant of an exponentially growing population. These rates are also determined for varying levels of drug concentrations. The dynamics of tumor cell death can be characterized using apoptosis assays such as Annexin V/propidium iodide (PI) fluorescence-activated cell sorting (FACS) assays [54], where cells with positive Annexin V staining are considered to be dead. Other assays for proliferation (e.g. KI-67 protein assay [31]) and death (e.g. Caspase activation assays [33]) can also be used to determine the respective rates.

In the computational approach implemented in ACESO, data from the viability and apoptosis assays (cell levels measured over time in absence or in presence of drug(s) concentrations) are used in a two steps procedure as follows:

(i) For each treatment condition (concentration level of one or two drugs) the corresponding first order rate constants of birth (b_i) and death (d_i) are estimated using equations 3.4 and 3.6.

(ii) **A**. In case of a exposure to a single drug, modeling the b_i and d_i estimates vs the full concentration range profiles provided the values of the drug effect parameters, as EC_50, B_max, and$h$ in equation 3.7 (see *Quick guide to equations and Assumptions*), although other model alternatives are available in ACESO (Weibull, Gompertz, linear models, etc.). The selection of the best model is based on the lowest Akaike Information Criterion (AIC) values [1].

B. The relationship between estimates of b_i or d_i and drug concentrations obtained from drug combination studies was established using generalized additive models (GAMs) [5] or locally weighted scatterplot smoothing regression (loess) [1], both implemented in ACESO and belonging to the class of nonparametric models. Thus, the b_i or d_i vs concentration profiles are described without assuming any mechanism of interaction which provides a high degree of flexibility in the fitting but lacks any biological interpretation.

Estimating the mutation rate from data is still an unsolved problem which has not been yet adequately addressed in literature. Thus for the simulations exercises from the Results section, very small values ranging between $10^{-7} - 10^{-9}$ for the mutation rates are arbitrarily defined.

Pharmacokinetics

In the case where patient specific PK data is available, user-defined or compartmental models already coded in ACESO using the *mrgsolve* model specification (one to three compartments considering intravenous and extravascular route of administration) are available to select the best model describing the concentration vs time curves. This functionality from ACESO is called the *PK curve estimator*. If data from different individuals are available, parameter estimates describing the median tendency of the data will be reported, as estimation of inter-individual variability is not supported in the current version of ACESO. Selection between models is based mainly on the minimum AIC value and inspection of the individual fits.

Assessing drug synergy/antagonism

Since the analysis of drug combination data using the nonparametric models described above does not provide any information on the type of interaction between compounds, ACESO helps to identify if the combined action of two drugs is synergistic, antagonistic or simply the drugs don't interact with each other.

In order to quantify the degree of synergy/antagonism between two compounds, the most common approach is to compare their measured combination effect to a null reference model of no interaction, i.e. the expected response assuming no interaction between the two compounds. If the combination response is greater than what is expected by the reference model, the combination is classified as synergistic, while antagonism is defined when the combination produces less than the expected effect.

There are several well-known conventional approaches that define different null models to assess drug synergy/antagonism. The Loewe Additivity model [42] is one of the most commonly used models to quantify a zero-interactive state for the combination of two drugs. This model is based on the assumption that a drug cannot interact with itself and defines synergy/antagonism as a combined inhibitory effect that is greater/lower than the sum of the individual effects of the drugs. Highest Single Agent (HSA), also known as Gaddum's non-interaction model [3], is another popular model which defines a "independent action" of the drugs when the predicted effect of a combination is that of the one most effective drug alone. According to this model, any combined effect stronger than the effect of a single drug is called 'synergism' and a weaker effect 'antagonism'. Additional information and equations of these models are described in Supplementary material.

All these methods represent a pure comparison between observed and no-interaction model response and they don't define a mechanistic model to quantify the interaction of the drugs. Even so, both reference models are implemented in ACESO and allow for the computation of a response surface which can then be used for the simulations of the evolutionary process when there is a clear certainty that the drugs do not interact with each other.

3.2.3 Model predictions: Simulation of the treatment effects

Once the birth and death first order rate constants for all the cell types are defined as a function of drug concentrations and the PK model(s) are selected, simulations of potential dosing regimens can be performed aiming to find the optimal therapeutic strategy that minimizes the expected number of total tumor cells remaining and/or the probability of developing resistance within a specified time frame.

Although the current version of ACESO doesn't support the estimation of inter-individual variability (IIV), it allows for the simulation of IIV in the model in the case where population PK models from the

literature or model parameters estimated using platforms based on nonlinear-mixed effects methods (e.g. NONMEM [3], MONOLIX [40]...) want to be used. Therefore, ACESO allows exploring the impact of interpatient pharmacokinetic variability and adherence to treatment (e.g., missed doses) on disease progression and dynamics of resistance.

3.3 Results

Figure 3.2 lists all the functionalities of ACESO that now will be deeply explored through different demonstrative case studies.

In the first case study, we demonstrate the use of ACESO to determine optimum combination dosing strategies using publicly available data from the HMS LINCS Database. In the second, a previously developed branching process model that evaluates the evolutionary processes driving the emergence of resistance in response to erlotinib in non-small cell lung cancer (NSCLC) is extended to account for CNS metastases. These case studies highlight the fact that quantitative knowledge of pharmacokinetic, drug interaction and evolutionary processes is essential for identifying best intervention strategies.

3.3.1 Case study 1: LINCS database

The HMS LINCS Database is a publicly available database that aims to collect and disseminate data which can be used to understand how human cells respond to different types of perturbation such as exposure to drugs, environmental conditions, and mutations [34].

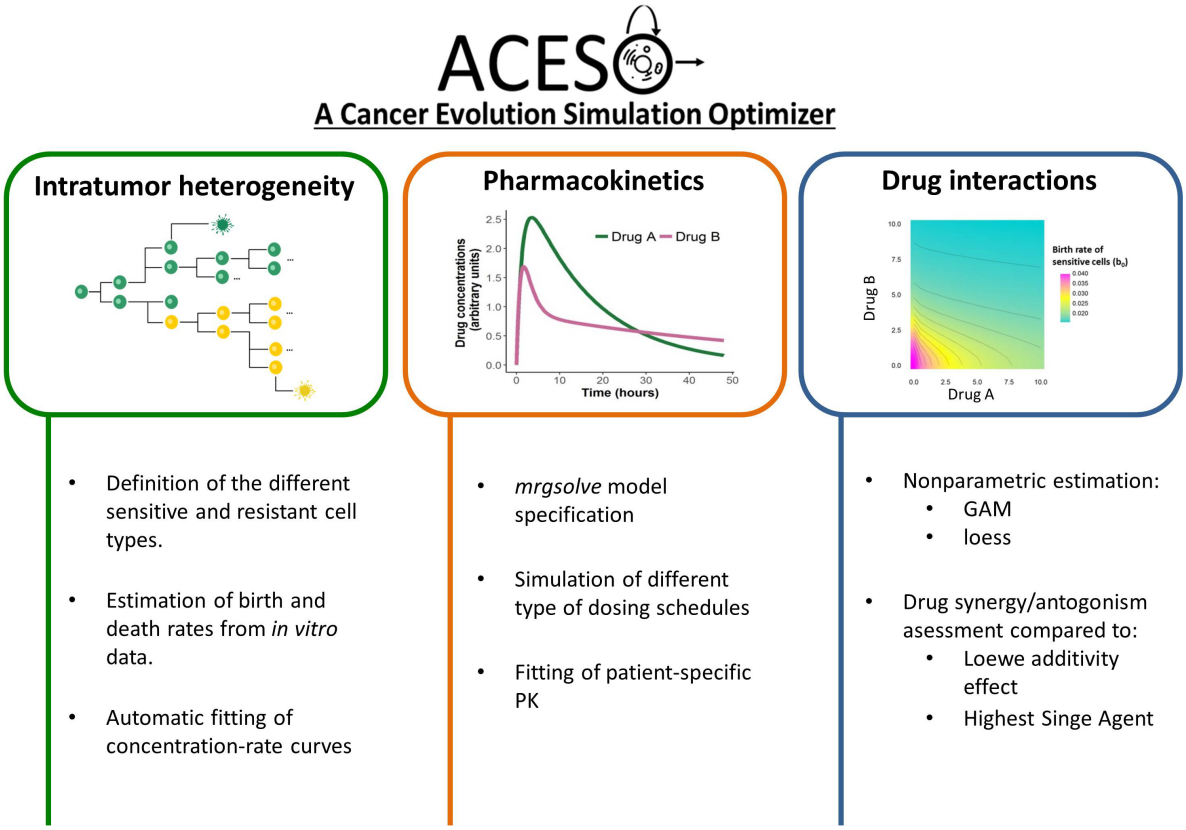


Fig. 3.2 Main functionalities of ACESO. PK refers to pharmacokinetics, GAM to Generalized Additive Models and loess to locally weighted scatterplot smoothing.

In the current evaluation cell viability data resulting from the exposure of BT-20 triple-negative breast cancer (TNBC) cell line to different concentration values of the targeted drugs alpelisib, neratinib, and trametinib used as single agents or in combination between (i) alpelisib and neratinib, or (ii) alpelisib and trametinib, were analyzed following the evolutionary process and the modeling and simulation steps described in Material and Methods in order to analyze the dynamics of heterogeneous TNBC tumor cell populations under combination treatment.
Briefly, in these experiments cells were exposed during 72 hours to constant concentrations levels of the drugs either as single agents or in the combinations described above. The reader is referred to the HMS LINCS page (http://lincs.hms.harvard.edu/db/) for more information regarding the experimental settings. The raw data used in the present analysis can be retrieved from the datasets identified as 20245, 20247 and 20259.

Regarding the drugs used in the experiments, neratinib is an irreversible tyrosine kinase inhibitor of HER1, HER2, and HER4 [5] which is currently being used for the extended adjuvant treatment of early stage HER2-positive breast cancer following adjuvant trastuzumab-based therapy. Alpelisib is a phosphatidylinositol-3-kinase (PI3K) inhibitor which is currently undergoing different clinical trials including a phase III study where the combination of alpelisib with fulvestrant (hormonal therapy) as a second-line therapy for patients with ER-positive/HER2-negative breast cancer whose tumours had PIK3CA mutations is being tested [52]. Finally, trametinib is a reversible inhibitor of mitogen-activated extracellular signal regulated kinase (MEK) activation [21] which has been approved by the FDA for the treatment of metastatic melanoma with BRAF V600E or V600K mutations.

Analysis of Alpelisib and Neratinib combination

We first analyzed the data from the combination of Alpelisib and Neratinib drugs. BT-20 cells sensitive to both drugs were defined as type 0 cells in our framework, whereas cells resistant to neratinib and alpelisib were defined as type 1 and type 2 cells respectively.

Dynamics of sensitive (type 0) cells

Figure 3.3A shows the relationship between viable cells measured 72 h after exposure and the different drug concentration levels, where it can be observed that both targeted drugs reduce in a concentration-dependent manner tumor cell viability and that greater reduction of cell viability is observed for the two drugs in combination compared to single drug effects. Moreover, the effect of alpelisib seems to be stronger at first sight, although we must note that the maximum concentration of neratinib used in this combination experiment ($1.92\mu M$) is much lower than the maximum concentration of alpelisib ($10\mu M$).

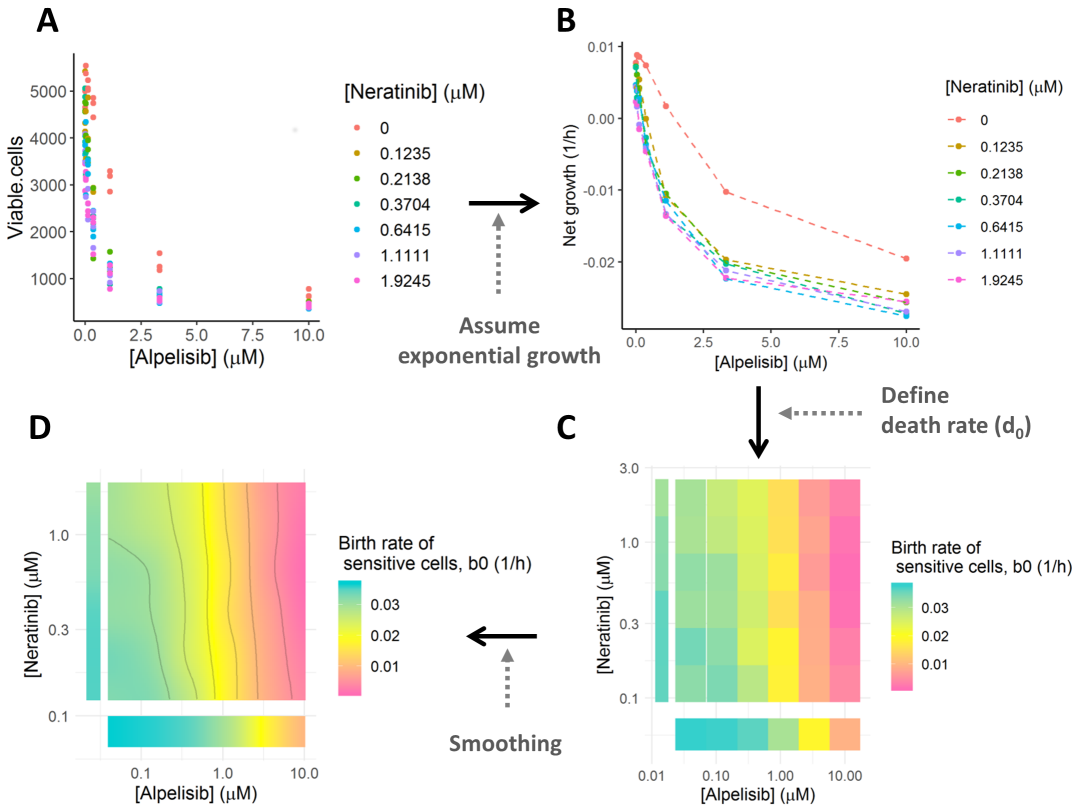


Fig. 3.3 Estimation of the net growth rate of sensitive cells using drug combination data. A) Viable cell counts under different drug concentrations of alpelisib and neratinib. B) Net growth rate of sensitive cells as a function of the different drug concentrations. C) Discrete values of the birth rate of sensitive cells under different drug concentrations. D) Smooth response surface of the sensitive cell birth rate values under different drug concentrations.

An exponential model was used to deterministically characterize the change in viable cells from 0 to 72 h:

$$N(t) = N_0 \cdot e^{\lambda_{0,j} \cdot t} \tag{3.10}$$

where N is the number of viable cells, N_0 is the number of viable cells at time 0 (2751 in this database) and $\lambda_{0,j}$ is the first order net growth rate parameter of type 0 cells corresponding to each treatment condition j (drug(s) and concentration level) reflecting the difference between birth and death first order rate constants and referred as λ_0 hereafter. The estimates of λ_0 ranged from $7.73x10^{-3}\,h^{-1}$ in absence of both drugs to $-2.55x10^{-2}\,h^{-1}$ in presence of 10 μM of alpelisib and 1.92 μM of neratinib. Figure 3.3B shows the net growth rate estimates vs concentration profiles.

As previously explained, the evolutionary model requires values for the cell birth, death and mutation rates. However, as there was no information regarding apoptosis assays available to estimate the first order death rate of type 0 cells (d_0) and assuming that these targeted therapies do not induce cell death

per se, a first order elimination rate constant of 0.028 h^{-1} was defined to ensure that all the values for the first order birth rates of type 0 cells (b_0) are positive regardless of the treatment conditions.

The values of b_0 are listed in the matrix shown in Figure 3.3C and Figure 3.3D shows the same matrix after using a smoothing function. These parameters were fitted nonparametrically using a GAM to generate the predicted surface (Figure 3.4A) required for the simulation of the evolutionary process.

Dynamics of drug resistant cells (type 1 and type 2) cells

Once drug resistant cells have mutated from drug sensitive cells they follow their own fate characterized by their corresponding bi and di. Due to the lack of information regarding drug-resistant cell lines in LINCS, we assumed that these cells had decreased birth and death rate parameters compared to their sensitive counterparts, and hence we divided the proliferation and death rate parameters obtained in the previous exercise for the parental BT-20 cell line by different arbitrary values depending on the drugs being analyzed (by two in the case of alpelisib resistant cells and by three for neratinib resistant cells).

A full resistance mechanism was considered in this example, that is, neratinib drug concentrations didn't have any influence on neratinib-resistant cell proliferation and therefore, these cells are only affected by the varying alpelisib concentrations as can be seen in the second panel of Figure 3.4A. The same is assumed regarding alpelisib-resistant cells (third panel in Figure 3.4A). Those profiles were characterized using the nonlinear models for single agents included in ACESO providing very good fits (an exponential decay model and a five parameter log-logistic function best fitted the first order birth rates of neratinib-resistant cells and alpelisib-resistant cells respectively).

In addition, a value for the second order mutation rate constant of 10^{-7} was defined for the cells resistant to one drug, and of 10^{-8} for cross-resistance (resistant to both drugs). Although a modification of these estimates would alter the results of the simulations slightly, it would not affect the relative comparison between dosing schedules.

Pharmacokinetic models

Pharmacokinetic parameters required to generate the concentration vs time profiles for alpelisib were obtained from a previous pharmacokinetic analysis published in literature [14]. In the case of neratenib, pharmacokinetic parameters were estimated from an analysis performed with the median concentration vs time data scanned from the original publication [36] using WebPlotDigitizer (https://apps.automeris.io/wpd/). Results from this analysis performed using the *PK curve estimator* from ACESO are shown in supplementary figure S3.2.

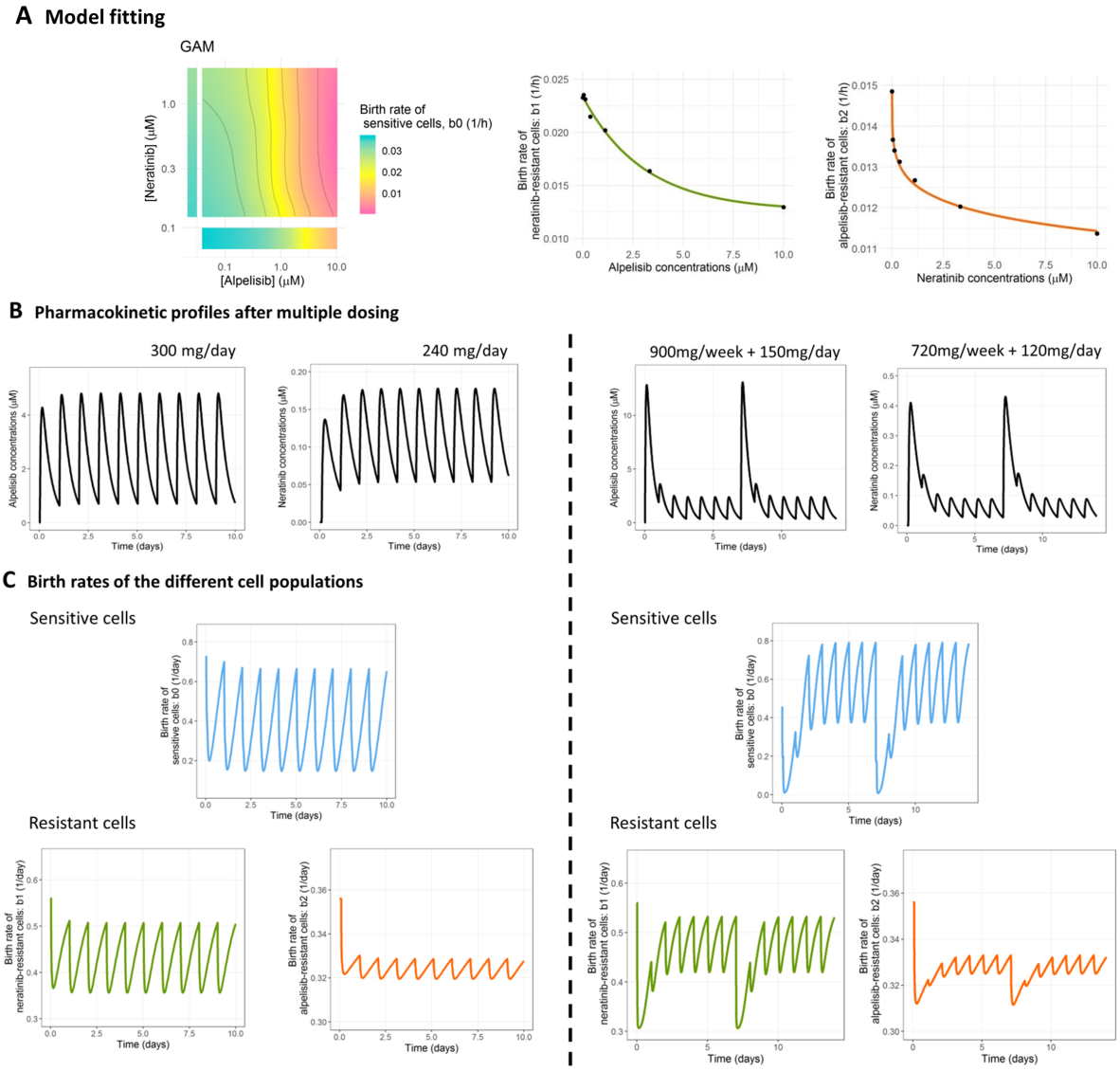


Fig. 3.4 A) Different fitted models for the data used in the case study of the combination of alpelisib and neratinib drugs: a GAM is fitted to describe the effect of the combination of the two drugs on the birth rate of sensitive cells (left), an exponential model best described the alpelisib concentration-response curve of neratinib-resistant cells (middle) and a five parameter log-logistic model best fitted the birth rate values under different concentrations of neratinib of alpelisib-resistant cells (right). B-C) Effect of the varying plasma drug concentrations over time (B) in the birth rate of the different sensitive and resistant cell types (C).

Alpelisib is in late stage clinical trials, where a standard dosing schedule of 300 mg/day orally is being tested when the drug is given in combination with other anticancer agents [45]. The recommended therapeutic oral dose for neratinib on the other hand, is 240 mg/day continuously for one year. Figure 3.4B shows the time course of alpelisib and neratinib drug concentrations after administration of two different dosing schedules (standard continuous daily dosing vs a high-dose pulse with continuous low-dose therapy) and how the birth rates of the different cell populations change over time as a

consequence of the time varying concentrations (Figure 3.4C).

In the next section, the differences caused by the treatment on the total number of the different cancer cell types and the probability of developing resistance calculated using the equations 3.1 and 3.2 were explored.

Simulation of the treatment effects

Having defined (i) the PK models, (ii) the response surface for the effect of the combination of the two drugs on the proliferation of the sensitive cells, (iii) the birth rates vs concentration curves and death rates of drug resistant cell types and (iv) the mutation rates, ACESO was used to simulate the impact of different dosing schedules on the dynamics of resistant cells assuming a population of 1 million sensitive cells and 1000 resistant cells to alpelisib and neratinib prior to treatment initiation. In addition to the standardized MTD schemas for alpelisib and neratinib, we investigated the dosing strategies listed in Figure 3.5A among others.

We first compared between sequential, alternating and simultaneous dosing strategies and concluded that simultaneously administering the drugs gave the best results (data not shown). Among the simultaneous dosing strategies tested, schedules including once-weekly pulses with resting periods resulted in the highest expected number of tumor cells after the treatment period. These results are caused by the fast clearance of the drugs that allow cancer cells to grow during treatment breaks as there are no drug concentrations decreasing the value of their birth rates. Increasing the weekly pulse frequency (twice-weekly, three times a week) did not improve the results (Figure 3.5B-C). For the case of alpelisib, combining a high-dose pulse of 900 mg with a maintenance low daily dosing of 200 mg for the rest of the week showed a greater effect, but proved to be a 12% worse than the daily dosing of 300 mg in the lessening of the number of alpelisib-sensitive cells (type 1 in the Figure 3.5C) and almost 5% worse in the effect caused to type 0 cells. By contrast, low neratinib concentrations have a considerable effect on the birth rate of type 2 cells and therefore the schedule involving a high-dose pulse of 720 mg/week with an additional daily dose of 120 mg for the remainder of the week was comparable to the approved schedule of 240 mg once a day (QD) (see simulation 1 and 7 of Figure 3.5). Thus, both the dosing schedules from simulation 1 and 7 gave the most satisfactory results. The decision of using one or the other could be based on the tolerability of each dosing regimen by the TNBC patients.

A

Dosing regimen	Alpelisib	Neratinib
1	300 mg QD	240 mg QD
2	300 mg QD	120 mg QD
3	300 mg QD	360 mg QOD
4	300 mg QD	1200 mg QW
5	300 mg QD	720 mg QW + 40 mg QD r.m.
6	300 mg QD	720 mg QW + 80 mg QD r.m.
7	300 mg QD	720 mg QW + 160 mg QD r.m.
8	525 mg QOD	240 mg QD
9	900 mg QW + 200 mg QD r.m.	240 mg QD
10	900 mg QW + 200 mg QD r.m.	720 mg QW + 160 mg QD r.m.

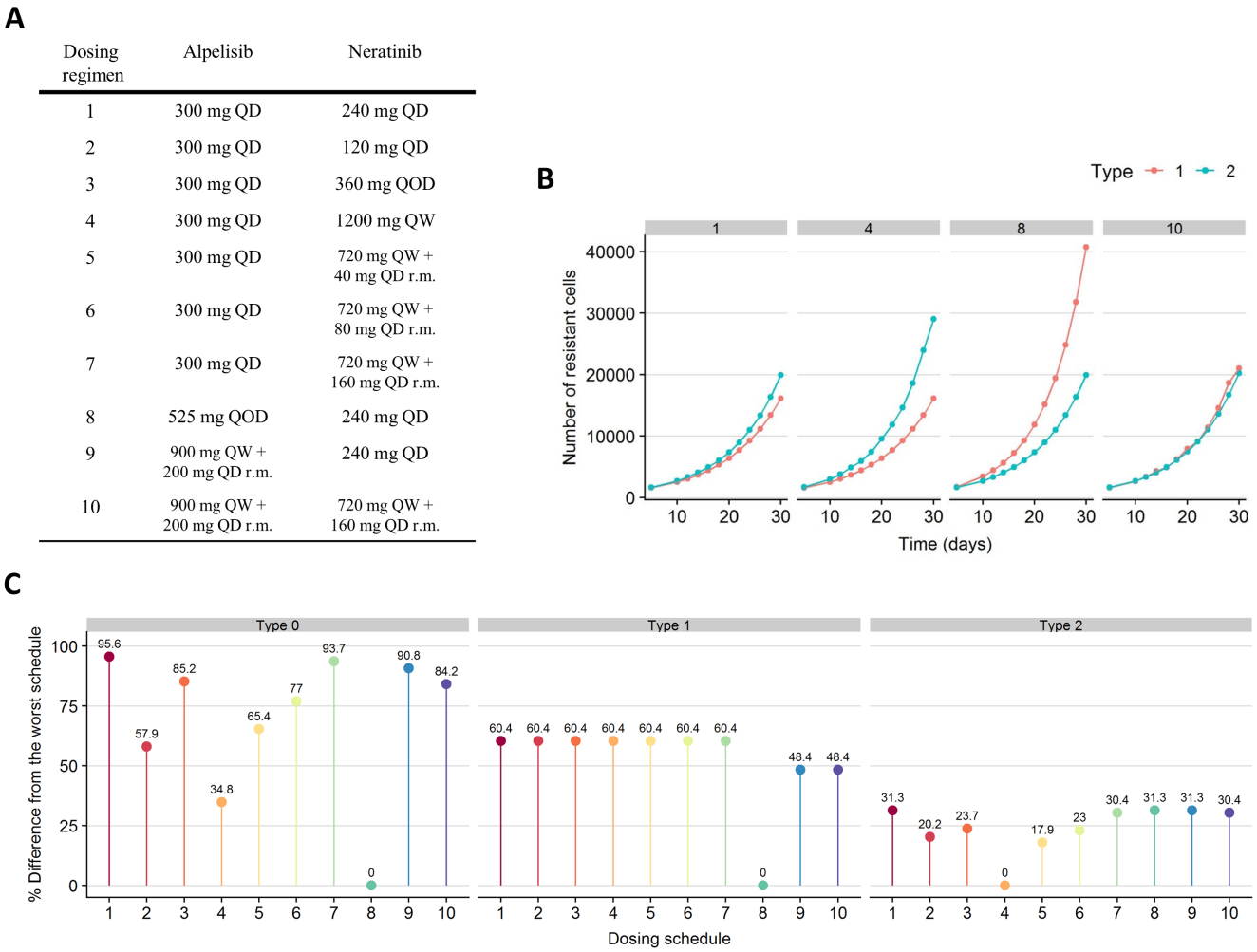


Fig. 3.5 Effects of the treatment on the total number of the different type of cancer cells. A) Different dosing schedules explored (QD: once a day, QW: once a week, and QOD: every other day). B) Effect of four simulated dosing schedules on the number of resistant cells over time. C) Comparisons of the outcomes of the different dosing schedules. The best regimen (least expected number of cancer cells at 30 days) has the highest bar, while the worst regimen has zero height.

Analysis of the Alpelisib and Trametinib combination

The previous example has shown several of the main capabilities of ACESO. In this second case study, the analysis is focused on the characterization of the type of interaction occurring in drug combination therapies, as well as comparing the magnitude of the effect elicited by different combinations (alpelisib and neratinib vs alpelisib and trametinib). The dynamics and impact of emerging resistant cells is, in this case, out of scope.

The procedure to calculate the first order net growth, birth and death rate parameters of the sensitive cells under different combinations of alpelisib and trametinib was analogous to the previous example.

Pharmacokinetic models

The pharmacokinetic model and parameters for trametinib were also extracted from literature [49]. In brief, the disposition properties of trametenib in plasma were described with a two-compartment model with linear elimination. Absorption after oral administration resulted to be time variant evolving from a slow to a fast absorption process in a short period of time (0.4 h). For simplicity, only the first order absorption rate constant equal to 2.05 h^{-1} was used. Plasma drug concentration vs time profiles of trametinib were then generated for the recommended schedule consisting on 2 mg orally once a day.

PK characteristics for alpelisib were already described in the previous example.

Characterizing the type of interaction

To determine whether the combination of alpelisib and trametinib elicits a greater reduction of the birth rates of type 0 cells than the expected under the assumption of additivity or from the effect of each drug alone, a visual comparison was performed between the response surface resulting from the non-parametric analysis of the combination data and the response surface generated from the generalized Loewe additivity or HSA models (see supplementary material for the specific equations and procedure).

In Figure 3.6A it can be appreciated that the effect caused by trametinib when combined with high alpelisib concentrations is negligible as the vertical isobole lines from the graph indicate that infinite concentration values of trametinib are required to achieve the same effect than the one produced by $\sim$ 1-10 μM alpelisib concentrations. In addition the three surfaces appear to be comparable suggesting lack of interaction. However, the profiles shown in Figure 3.6B where the expected number of sensitive cells is simulated over time (based on the results from the non-parametric analysis, Loewe additivity assumption and HSA model), reveal that at least some therapeutic benefit is seen after three

weeks of combination treatment at the recommended dosing (although we note that no statistical test was applied to know if this result was significant or not).

An equivalent analysis was performed for the alpelisib and neratinib combinations where the profiles shown in Figure 3.6D indicate a higher synergistic interaction. Consequently, between the two options explored in this case study, neratinib would be the most suitable choice to combine with alpelisib in future clinical trials.

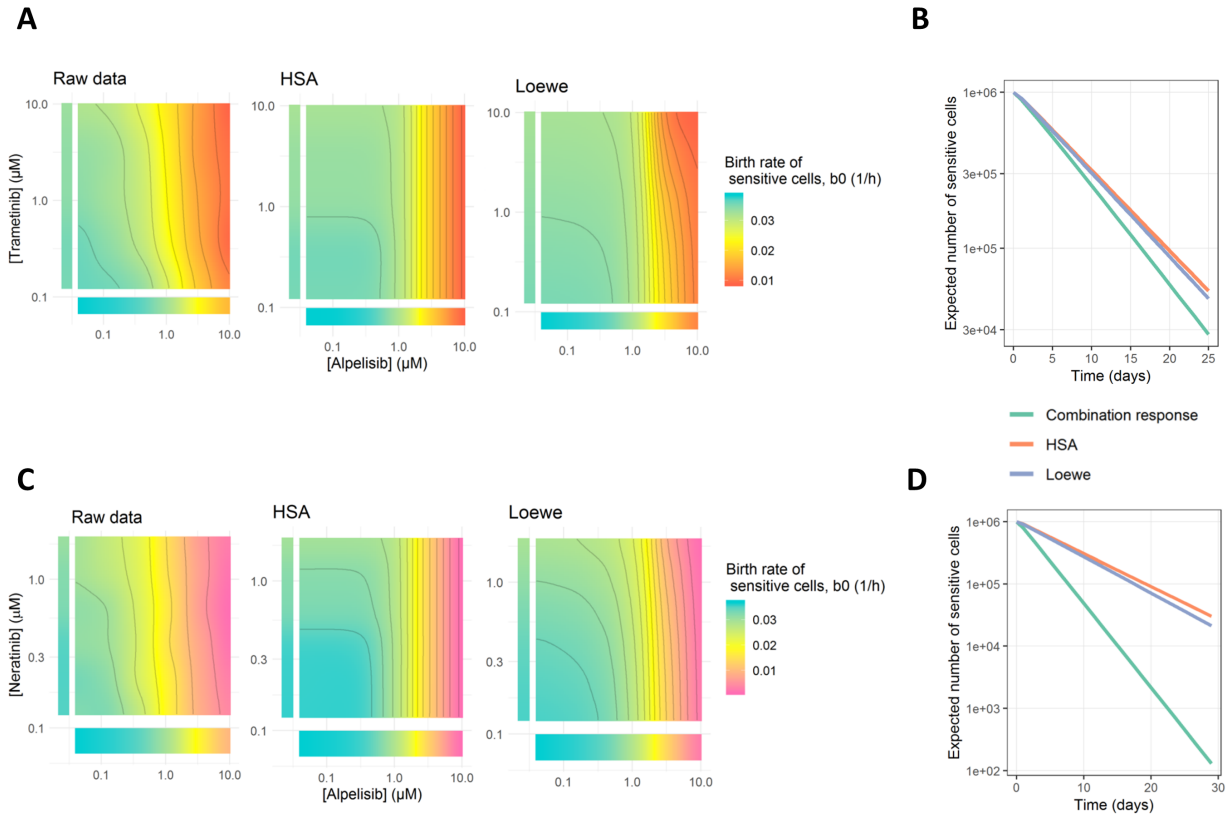


Fig. 3.6 A, C) Response surface analysis. Smooth response surface of type 0 cell birth rates as a function of alpelisib and trametinib concentrations (A) or alpelisib and neratinib concentrations (C) and its comparison to the surface obtained under Generalized Loewe additivity model and Highest-Single Agent hypothesis. B) Simulation of a simultaneous dosing strategy involving 300 mg QD alpelisib and 2 mg QD trametinib for 25 days and its effect on the total number of type 0 cells using the different response surfaces for type 0 cell birth rates showed in A. D) Simulation of a simultaneous dosing strategy involving 300 mg QD alpelisib and 240 mg QD neratinib for 30 days and its effect on the total number of type 0 cells using the different response surfaces for type 0 cell birth rates showed in C.

3.3.2 Case study 2: Erlotinib

In this second case study, we focused on an evolutionary cancer model from [8, 15] used to identify novel treatment administration strategies for non-small cell lung cancer (NSCLC).

NSCLCs that harbor mutations within the epidermal growth factor receptor (EGFR) gene are sensitive to the tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib [44, 50]. Unfortunately, all patients treated with these drugs acquire resistance, most commonly as a result of a secondary mutation within EGFR (T790M mutation). In addition, approximately one third of the patients develop CNS metastases after initial response to EGFR TKIs [48, 26, 29].

In this case study we extended the original two-type branching process from the previous publications to account for CNS metastases and to show the flexibility of ACESO to handle different variants of clonal evolution. As it has been suggested in previous works, the metastatized cells retain EGFR TKI sensitivity if sufficient drug concentrations can be achieved in brain parenchyma for brain metastases or in cerebrospinal fluid (CSF) for leptomeningeal metastases [26, 29]. Therefore in this model, three type of cells were considered (see Figure 3.7 for a schematic representation): primary tumor drug sensitive cells, primary tumor drug resistant cells and drug sensitive metastatized cells (drug sensitive cells that has disseminated to a new site and that have the same proliferation and death rates as the original cells).

Estimation of the birth and death rates of drug sensitive and resistant cells

Because both drugs were developed to target wild-type EGFR, we hypothesized that current dosing schedules were not optimized for mutant EGFR or to prevent resistance. To investigate this hypothesis, we used the data from previously developed isogenic erlotinib-sensitive and erlotinib-resistant pair of PC-9 cell lines that mimic the behavior of human tumors [8]. In order to estimate net (birth minus death) growth rates of erlotinib-sensitive and erlotinib-resistant cells, the total number of viable cells in culture was counted at 48, 60, and 72 hours in the presence and absence of the drug (see Figure 3.8 for the raw values). The subsequent steps to calculate the net growth rate parameter from the data assuming an exponential growth of the cells has been already explained in the previous case study. For this exercise, the data corresponding to time 0 was not used to calculate the net growth rate parameter. Death rates were estimated using equation 3.6 and the results of Annexin V/PI FACS assay where the fraction of death cells was determined at 48, 60, and 72 hours and for different concentrations of erlotinib (Figure 3.8 bottom). For the resulting birth and death rates and more details about the experimental conditions, see the work from Chmielecki et al. [8].

We fitted the concentration-response curves for the birth and death rates of each cell type with ACESO (data not shown). The selected model for erlotinib-sensitive cells was a five parameter logistic function

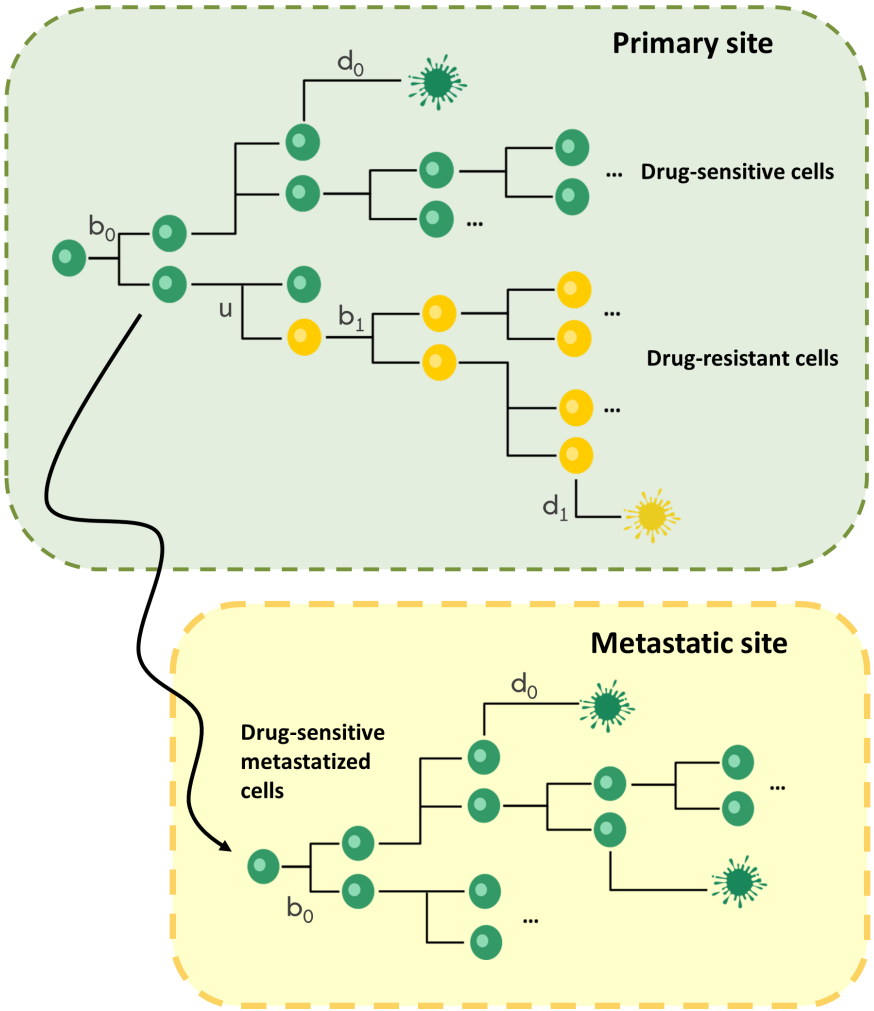


Fig. 3.7 Schematic representation of the stochastic model proposed in case study 2, where erlotinib-sensitive cell (green), proliferates at rate $b_0(t)$, dies at rate $d_0(t)$ and can accumulate mutations at rate u per cell division to give birth to the erlotinib-resistant cell type (yellow), which proliferates at rate $b_1(t)$ and dies at rate $d_1(t)$. Erlotinib-sensitive cell is also able to migrate to other sites (to CNS in this example) and continue growing at the same rates as the original cell type.

(see equation 3.8) with a cell intrinsic birth rate of 0.04 h^{-1}, a B_{max} of 0.003 h^{-1}, a EC_{50} value of 0.76 μM , a Hill coefficient of 1.5 and an asymmetry factor of 8.26. In contrast to previous case studies, a model was also fitted for the small influence of erlotinib in the growth inhibition curve of drug- resistant cells instead of completely neglecting the erlotinib influence on the resistant variant. Although it was not the best fitted model (the model with lower AIC), a linear equation with an intercept of 0.034 h^{-1} and a slope parameter of -0.0005 ($\mu M \cdot h^{-1}$) was used to model the birth rates of the resistant PC-9 cell line as in the original article. The death rates obtained from the Annexin V/PI FACS counts for both cell types were almost identical for the different drug concentrations used in the experiments, and hence we defined a constant death rate of 0.12 day^{-1} and 0.06 day^{-1} for the

drug sensitive and resistant cell populations respectively.

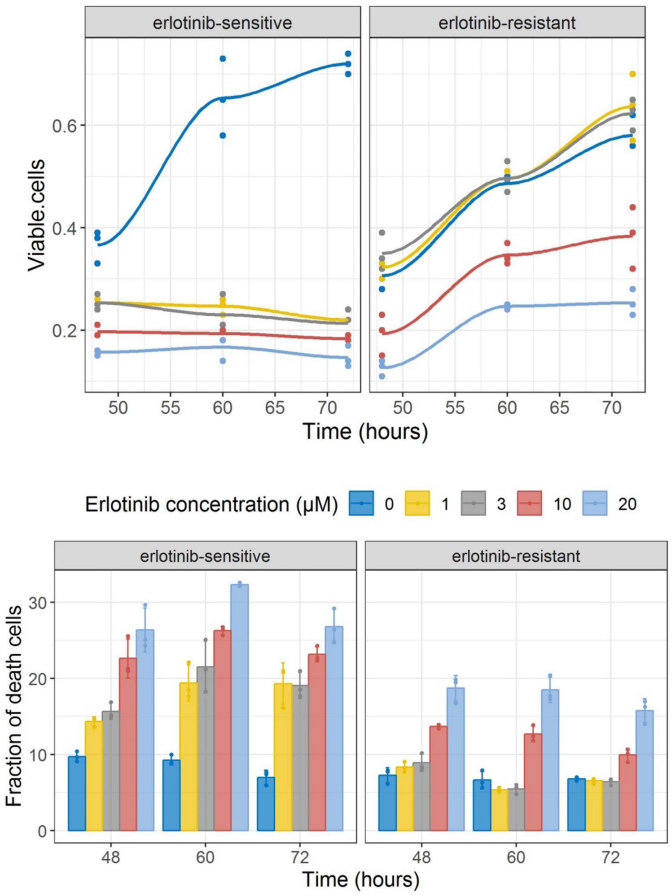


Fig. 3.8 (Top)Total number of viable cells (drug sensitive and resistant) over time and for different concentrations of erlotinib. Points represent the raw observations and solid lines indicate the smooth line obtained for the three replicates for each erlotinib concentration (Bottom). Fraction of death cells (%) over time and for different erlotinib concentrations. Points represent raw values, bars indicate the median value of the three replicates for each drug concentration tested and error bars indicate the range of the raw values.

Pharmacokinetic model

The erlotinib plasma drug concentration vs time profiles were generated using the model and the corresponding model parameters obtained from a pharmacokinetic analysis performed with data from 1047 patients with solid tumors [43], instead of using simpler equations for the PK obtained from scarce data points from the literature as in the publication from [15].

Simulations represented in Figure 3.9A were performed considering the administration of a 1600 mg dose of erlotinib once a week over one month of treatment and showed the differences in exposure obtained when using each of the models. The estimated drug clearance from Foo et al. [15] is faster

than the parameter reported in the population model from Lu and coauthors [43], which causes the cells to resume proliferation as there are no drug concentrations to inhibit cell growth during treatment breaks. Indeed, these differences had a huge impact on the risk of developing resistance as can be seen in Figure 3.9B. This example highlights the importance of properly characterizing the PK of the drugs being analyzed as they can have a high influence in selecting optimum dosing strategies.

On the other hand, previous studies had demonstrated that the concentration of erlotinib measured in cerebrospinal fluid (CSF) during standard daily dosing of 150 mg is too low (5% of that in plasma according to Togashi et al. [53]) to stop EGFR-mutant NSCLC cell proliferation, supporting the hypothesis of poor erlotinib delivery into the CNS rather than the development of acquired drug resistance and converting CNS a common site of disease progression. Thus we assumed a linear pharmacokinetic process (dose independent) with a partition coefficient of 0.05 to simulate erlotinib concentrations in the metastatic site.

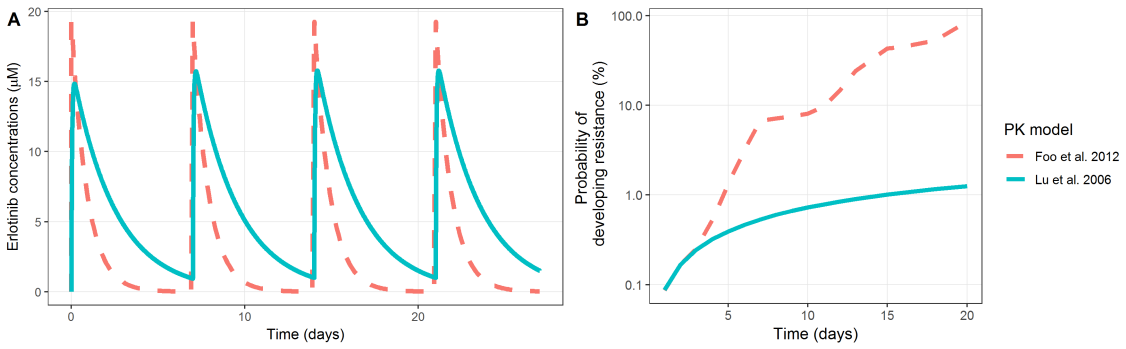


Fig. 3.9 Effect of two different pharmacokinetic models for erlotinib on the probability of developing resistance. The pharmacokinetic models described in 1) Lu et al. 2006 [43] and 2) Foo et al. 2012 [15] were coded in ACESO and a dosing schedule of 1600mg once a week was simulated to explore the effects of the resulting erlotinib concentrations (A) on the risk of developing resistance for a period of 20 days (B).

Simulation of the treatment effects

The current analysis explores whether a higher dose of erlotinib provides an increased concentration in CNS that may prove to be more efficient in metastatic CNS cancers while preventing the development of resistance at the primary site. Based on the above considerations, the standard oral daily dosing schedule of 150 mg was compared to a high-dose pulse of 1600 mg once a week and four different combinations of low-dose continuous and high-dose pulsed strategies (see Figure 10 to know the particular dosing regimens simulated) assuming one month of treatment. Additional simulations increasing the daily dosing for erlotinib were not tried because daily doses above 200 mg induce unacceptable toxicity [46]. In contrast, weekly high-dose pulses up to 2000 mg were tolerated by NSCLC patients despite of persistent nausea reported in patients receiving higher doses than 1200 mg [9].

Simulations were performed in the case of (i) absence of pre-existing tumor cells at the metastatic site at the time of tumor diagnosis, and (ii) CNS established metastases at the moment of the primary tumor diagnosis, assuming no pre-existing resistance in both cases. A mutation rate of 10^{-8} per erlotinib-sensitive cell division was defined and a dissemination rate to CSF of 10^{-9}.

The resulting number of the different cancer cell populations is shown in Figure 3.10. Variations in the number of sensitive cells in the primary tumor site from one schedule to the next was minimal as well as the probability of developing resistance over time (data not shown). However, for erlotinib-resistance cells and especially for metastatized erlotinib-sensitive cells using high-dose pulses considerably improved the results.

ACESO also allowed the estimation of the probability of dissemination to CNS (as the equation for the probability of developing resistance can be customized for other conditions), which was also very similar for all the dosing schedules simulated with a value of approximately 1.38% for a treatment period of 25 days. Clearly, the task of finding the optimal dosing regimen could have been further complicated by allowing acquired resistance to arise in the metastatized cell population defined in our

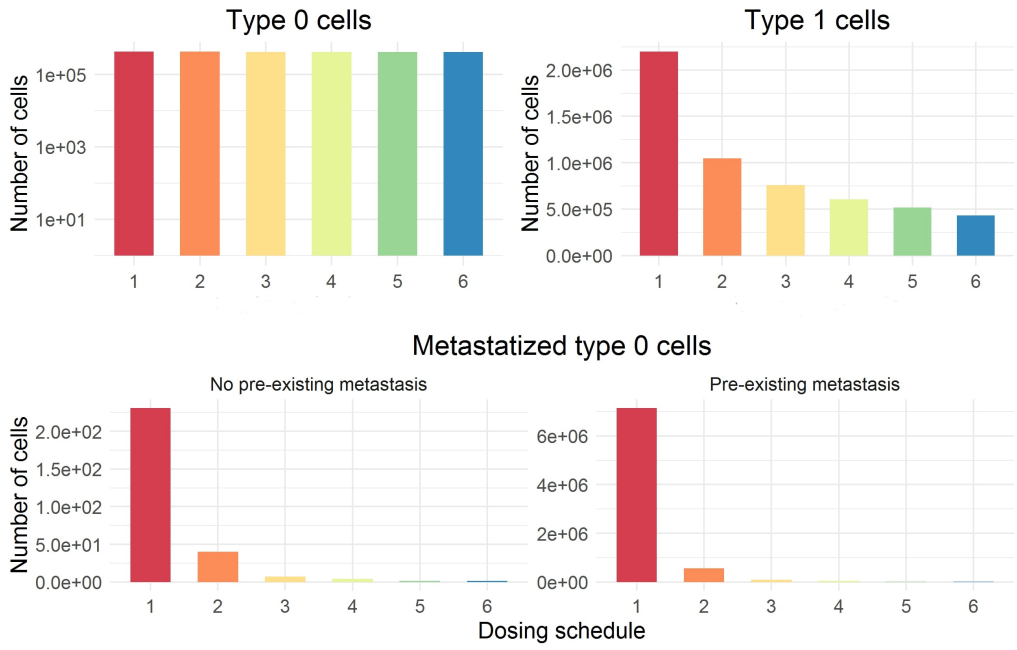


Fig. 3.10 Effect of different dosing regimens in the number of cancer cells at 30 days. Erlotinib dosing schedules used in the simulations: 1) standard daily dosing of 150mg, 2) high-dose pulse of 1600mg once a week, 3) high-dose pulse of weekly 1600mg plus 50mg/day the rest of the week, 4) high-dose pulse of weekly 1800mg plus 50mg/day the rest of the week, 5) high-dose pulse of 1000mg day 1 and 2 followed by 100mg/day the rest of the week and 6) high-dose pulse of 1200mg day 1 and 2 followed by 50mg/day the rest of the week. Type 0 cells refer to primary tumor drug sensitive cells, type 1 cells to primary tumor drug resistant cells and type 2 cells to drug sensitive metastatized cells. For the case of pre-existing metastasis before treatment initiation, an initial number of 50 cells were assumed. The initial number of type 0 cells was set to 10^6 and we defined a erlotinib penetration to CNS of 5%.

computational model.

We conclude that a combined low-dose continuous and high-dose pulsed erlotinib schedule was the most successful dosing strategy at preventing progression in patients with CNS metastases but did not show significantly delayed emergence of resistance due to the T790M EGFR mutation (although once resistant cells arise, these schedules also demonstrated to be more effective killing the resistant cell population).

3.4 Discussion

In this work, we have built a multi-scale computational framework in R called ACESO (A Cancer Evolution Simulation Optimizer) which incorporates a continuous time multitype branching process model with concentration-dependent birth, death and mutation rates to investigate the evolution of resistant clones to anti-cancer therapies. The emergence of drug resistance represents a significant obstacle to the successful control of tumors since it contributes to drug failure and poor prognosis. The use of mathematical models to search for optimized treatment schedules that delay the appearance of this resistance in cancer cells is part of a growing effort to improve clinical trial design for cancer patients. Here we provide an open-source tool to contribute to this search by exploring the dynamic evolution of heterogeneous tumor cell populations while taking pharmacokinetic and drug interaction effects into account.

Literature is plenty of examples showing different models to describe drug pharmacokinetics, tumor progression and drug effects. There are also tools that allow developing and fitting computational models to cancer data from different sources and nature (i.e., biomarkers, tumor volume, etc.). However, such ambitious enterprise usually requires the use of different platforms (for parameter estimation, model simulation or numerical and graphical diagnostics), delaying the application of integrative quantitative analysis of cancer data. The objective of the current work was to integrate in a single computational framework the tools required to characterize all the process mentioned above, becoming, to the best of our knowledge, the first R package allowing the merge of deterministic and stochastic approaches, and paying special attention to the incorporation of pharmacokinetic processes and their influence in the growth kinetics of branching process models with the goal of finding optimized dosing strategies.

In order to simulate a wide range of PK models and dosing schedules, ACESO incorporates the mrgsolve package and provides the codes of the most commonly used pharmacokinetic models to ease the degree of competency needed to perform simulations based on ordinary differential equation systems. These PK models can then be further extended to increase their predictive value, for instance, by including the effect of patient-specific covariates. As an example, the PK model for erlotinib from [43] included the smoking status as a covariate in the parameter for the drug clearance and therefore additional simulations could have been easily performed to investigate the differences in the emergence of resistance found between smokers and non-smokers. These variables and complex dosing schedules are now easily simulated in ACESO, which provide an added value to the framework.

The application of our method to individual cancer types requires the availability of data on cell growth and death during different clinically tolerated concentrations of the drugs in order to parameterize the proliferation of sensitive and resistant cancer cell populations. ACESO uses of *in vitro* growth kinetics for the sensitive and resistant cancer cell populations, because of the difficulty of obtaining

similar concentration-response curves from *in vivo* settings. This fact complicates the predictions of realistic time scales for the emergence of resistance because growth kinetics *in vitro* occurs in a time scale orders of magnitude faster than *in vivo*. The comparisons between dosing schedules will always be valid as long as the relative differences in growth rates are constant between *in vivo* and *in vitro* situations or an *in vitro-in vivo* correlation model is included in the framework. Nevertheless, our approach enables the incorporation of a detailed set of measurements of in vitro assays for a wide range of drug concentrations, which in turn allows the model to predict the resistance dynamics under a variety of treatment protocols.

In this work we analyzed the freely available data from HMS LINCS database. The results of this analysis showed that when the drugs have fast elimination rates weekly high-dose pulses won't enhance the killing of cancer cells. Besides that, when the inhibitory effect caused by low and high drug concentrations is similar, as in the case of neratinib, the differences found between the standard daily dosing schedule and the weekly high dose pulse combined with a continuous low dose dosing regimen can be very small. However, if at least one of the drugs has a slow clearance and is able to accumulate over time, variations in the dosing regimens could induce significant differences in the total number of cancer cells at the end of the treatment period or in the probability of developing resistance [7]. In addition, the strong synergism found in alpelisib and neratinib showed that the drugs work best when dosed simultaneously and that the combination of these two drugs will produce a greater decrease in the number of drug-sensitive cells than the combination of alpelisib with trametinib. We must note however that the results of the first case study should be regarded carefully due to the lack of data regarding apoptosis assays and drug resistant cell lines. Still, the important aspect of this example is that it highlights the ability of ACESO to integrate all these models and parameters and generate plots (e.g. Figure 3.5) that allow a quick diagnosis and comparison of different dosing schedules.

In the second case study on the other hand, we showed that the stochastic process included in ACESO was previously applied to a population of EGFR-mutant cancer cells with the goal of rationally identifying optimum treatment administration schedules for NSCLC patients [8, 15]. In those publications, the growth kinetics for drug-sensitive and drug-resistant EGFR-mutant cells during therapy was determined and PK processes were incorporated into the model. This modeling framework enabled to identify a low-dose/high-dose pulse erlotinib administration schedule as the predicted best intervention strategy for stage IV NSCLC. This strategy was then tested in a prospective clinical trial [56] at the Memorial Sloan-Kettering Cancer Center (http://clinicaltrials.gov/show/NCT01967095), establishing the maximally tolerated dose as 1200mg for two days a week and 50mg for the remainder of the week. Although the study was not powered to demonstrate a significant delay of the emergence of T790M-driven resistance due to larger than expected variability in patient PK, a significant reduction in the rate of progression due to CNS disease was observed (0% versus up to 33% given historical control). In this work, we used the information of those publications and tried to reproduce their results

by expanding the model developed in [15] to include a population of drug-sensitive metastatizable cells that are able to migrate to the CNS. We also replaced their mathematical model for erlotinib by the population PK model developed in [43]. The resulting pulse dosing allowed for increased CNS penetration and the continued daily dosing controlled the progression of the disease. Due to the satisfactory results of the Memorial Sloan-Kettering Cancer Center clinical trial, this regimen is now being studied in a cohort of EGFR-mutant lung cancer patients with untreated CNS metastases demonstrating the tremendous anticipated translational potential of this approach.

ACESO is limited to two interacting drugs for now. Calculating the effect of each treatment on the birth or death rate of the different cell populations is a very complex task when more than two drugs are taken into account because of the difficulty in discerning the effect of each drug plus the synergistic/antagonistic interaction between them. In addition, the fact of using nonparametric models considerably increased the computation time needed for the simulations, although this problem could have been avoided with the incorporation of user-defined drug interaction models.

We are aware that the model integrated in ACESO is a simplified representation of the tumorigenic process as it only assumes exponential cell growth and no competing interactions between the cellular populations as this kind of interactions are hardly parameterized using in vitro data. In addition, spatial effects are not taken into account and therefore this platform won't be able to accurately characterize tumor evolution driven by microenvironmental determinants for now, like in the case of glioblastoma and its radiation response which greatly depends on the localization of the cell within the tumor [30]. Despite these facts, the flexibility of ACESO opens the possibility of further extending the underlying stochastic model to include more complicated processes that may play an important role in the development of resistance mechanisms.

In summary, the mathematical framework presented in this work is a general description of the evolutionary processes driving tumorigenesis, the emergence of resistance, and treatment response for oncogene-driven cancers which can be used to identify novel administration schedules of single-agent or combination treatment strategies that maximize efficacy and minimize the chance of resistance. Using ACESO, the different strategies can also be tested for robustness due to variability in pharmacokinetic parameters among patients, variable growth and death rates of sensitive and resistant cells as well as different compositions of the tumor at the start of therapy. We now propose to perform advanced validation of our methodology for other cancer types and for therapeutic agents at the forefront of clinical development.

References

[1] Akaike, H. (1987). Factor analysis and AIC. In *Selected Papers of Hirotugu Akaike*, pages 371–386. Springer.

[2] Baron, K. T., Hindmarsh, A. C., Petzold, L. R., Gillespie, B., and Margossian, C. (2017). Pastoor, d. mrgsolve: Simulate from ODE-Based population PK/PD and systems pharmacology models. metrum research group.

[3] Beal, S. L., Sheiner, L. B., Boeckmann, A. J., and Bauer, R. J. (2015). NONMEM user's guides (1989–2015). *Ellicott City: Icon Development Solutions.*.

[4] Beerenwinkel, N., Schwarz, R. F., Gerstung, M., and Markowetz, F. (2015). Cancer evolution: mathematical models and computational inference. *Syst. Biol.*, 64(1):e1–25.

[5] Bose, P. and Ozer, H. (2009). Neratinib: an oral, irreversible dual EGFR/HER2 inhibitor for breast and non-small cell lung cancer. *Expert Opin. Investig. Drugs*, 18(11):1735–1751.

[6] Bozic, I., Reiter, J. G., Allen, B., Antal, T., Chatterjee, K., Shah, P., Moon, Y. S., Yaqubie, A., Kelly, N., Le, D. T., Lipson, E. J., Chapman, P. B., Diaz, Jr, L. A., Vogelstein, B., and Nowak, M. A. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife*, 2:e00747.

[7] Chakrabarti, S. and Michor, F. (2017). Pharmacokinetics and drug interactions determine optimum combination strategies in computational models of cancer evolution. *Cancer Res.*, 77(14):3908–3921.

[8] Chmielecki, J., Foo, J., Oxnard, G. R., Hutchinson, K., Ohashi, K., Somwar, R., Wang, L., Amato, K. R., Arcila, M., Sos, M. L., Socci, N. D., Viale, A., de Stanchina, E., Ginsberg, M. S., Thomas, R. K., Kris, M. G., Inoue, A., Ladanyi, M., Miller, V. A., Michor, F., and Pao, W. (2011). Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Sci. Transl. Med.*, 3(90):90ra59.

[9] Clarke, J. L., Pao, W., Wu, N., Miller, V. A., and Lassman, A. B. (2010). High dose weekly erlotinib achieves therapeutic concentrations in CSF and is effective in leptomeningeal metastases from epidermal growth factor receptor mutant lung cancer. *J. Neurooncol.*, 99(2):283–286.

[1] Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991). Local regression models. chapter 8 of statistical models in S (edited by JM chambers and TJ hastie), 309-376.

[11] Coldman, A. J. and Goldie, J. H. (1983). A model for the resistance of tumor cells to cancer chemotherapeutic agents. *Math. Biosci.*, 65(2):291–307.

[12] Coldman, A. J. and Goldie, J. H. (1986). A stochastic model for the origin and treatment of tumors containing drug-resistant cells. *Bull. Math. Biol.*, 48(3-4):279–292.

[13] Cory, A. H., Owen, T. C., Barltrop, J. A., and Cory, J. G. (1991). Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun.*, 3(7):207–212.

[14] De Buck, S. S., Jakab, A., Boehm, M., Bootle, D., Juric, D., Quadt, C., and Goggin, T. K. (2014). Population pharmacokinetics and pharmacodynamics of BYL719, a phosphoinositide 3-kinase antagonist, in adult patients with advanced solid malignancies. *Br. J. Clin. Pharmacol.*, 78(3):543–555.

[15] Foo, J., Chmielecki, J., Pao, W., and Michor, F. (2012). Effects of pharmacokinetic processes and varied dosing schedules on the dynamics of acquired resistance to erlotinib in EGFR-mutant lung cancer. *J. Thorac. Oncol.*, 7(10):1583–1593.

[16] Foo, J. and Michor, F. (2009). Evolution of resistance to targeted anti-cancer therapies during continuous and pulsed administration strategies. *PLoS Comput. Biol.*, 5(11):e1000557.

[17] Foo, J. and Michor, F. (2010). Evolution of resistance to anti-cancer therapy during general dosing schedules. *J. Theor. Biol.*, 263(2):179–188.

[18] Foo, J. and Michor, F. (2014). Evolution of acquired resistance to anti-cancer therapy. *J. Theor. Biol.*, 355:10–20.

[19] Gadagkar, S. R. and Call, G. B. (2015). Computational tools for fitting the hill equation to dose-response curves. *J. Pharmacol. Toxicol. Methods*, 71:68–76.

[3] Gaddum, J. H. (1985). *Gaddum's Pharmacology*. Oxford University Press, USA.

[21] Gilmartin, A. G., Bleam, M. R., Groy, A., Moss, K. G., Minthorn, E. A., Kulkarni, S. G., Rominger, C. M., Erskine, S., Fisher, K. E., Yang, J., Zappacosta, F., Annan, R., Sutton, D., and Laquerre, S. G. (2011). GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. *Clin. Cancer Res.*, 17(5):989–1000.

[22] Goldie, J. H. and Coldman, A. J. (1979). A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. *Cancer Treat. Rep.*, 63(11-12):1727–1733.

[23] Goldie, J. H. and Coldman, A. J. (1983). Quantitative model for multiple levels of drug resistance in clinical tumors. *Cancer Treat. Rep.*, 67(10):923–931.

[24] Greaves, M. and Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, 481(7381):306–313.

[25] Groenendijk, F. H. and Bernards, R. (2014). Drug resistance to targeted therapies: déjà vu all over again. *Mol. Oncol.*, 8(6):1067–1083.

[26] Grommes, C., Oxnard, G. R., Kris, M. G., Miller, V. A., Pao, W., Holodny, A. I., Clarke, J. L., and Lassman, A. B. (2011). "pulsatile" high-dose weekly erlotinib for CNS metastases from EGFR mutant non-small cell lung cancer. *Neuro. Oncol.*, 13(12):1364–1369.

[27] Haccou, P., Haccou, P., Jagers, P., Vatutin, V. A., and . Vatutin, V. A. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.

[5] Hastie, T. J. (2017). Generalized additive models. In *Statistical models in S*, pages 249–307. Routledge.

[29] Hayat, M. A. (2014). *Brain Metastases from Primary Tumors Volume 1: Epidemiology, Biology, and Therapy*. Academic Press.

[30] Heddleston, J. M., Li, Z., McLendon, R. E., Hjelmeland, A. B., and Rich, J. N. (2009). The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell Cycle*, 8(20):3274–3284.

[31] Hogan, T., Yates, A., and Seddon, B. (2017). Analysing temporal dynamics of T cell division in vivo using ki67 and BrdU co-labelling by flow cytometry. *BIO-PROTOCOL*, 7(24).

[32] Iwasa, Y., Nowak, M. A., and Michor, F. (2006). Evolution of resistance during clonal expansion. *Genetics*, 172(4):2557–2566.

[33] Kaufmann, S. H., Lee, S.-H., Meng, X. W., Loegering, D. A., Kottke, T. J., Henzing, A. J., Ruchaud, S., Samejima, K., and Earnshaw, W. C. (2008). Apoptosis-associated caspase activation assays. *Methods*, 44(3):262–272.

[34] Keenan, A. B., Jenkins, S. L., Jagodnik, K. M., Koplev, S., He, E., Torre, D., Wang, Z., Dohlman, A. B., Silverstein, M. C., Lachmann, A., Kuleshov, M. V., Ma'ayan, A., Stathias, V., et. al. (2018). The library of integrated Network-Based cellular signatures NIH program: System-Level cataloging of human cells response to perturbations. *Cell Syst*, 6(1):13–24.

[35] Kendall, D. G. (1960). Birth-and-death processes, and the theory of carcinogenesis. *Biometrika*, 47(1/2):13–21.

[36] Keyvanjah, K., DiPrimeo, D., Li, A., Obaidi, M., Swearingen, D., and Wong, A. (2017). Pharmacokinetics of neratinib during coadministration with lansoprazole in healthy subjects. *Br. J. Clin. Pharmacol.*, 83(3):554–561.

[37] Kimmel, M. and Axelrod, D. E. (2002). Branching processes in biology (interdisciplinary applied mathematics).

[38] Komarova, N. (2006). Stochastic modeling of drug resistance in cancer. *J. Theor. Biol.*, 239(3):351–366.

[39] Komarova, N. L. and Wodarz, D. (2005). Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl. Acad. Sci. U. S. A.*, 102(27):9714–9719.

[40] Lavielle, M. (2005). MONOLIX (MOdèles NOn LInéaires à effets mixtes). *MONOLIX group, Orsay, France*.

[41] Liu, L. L., Li, F., Pao, W., and Michor, F. (2015). Dose-Dependent mutation rates determine optimum erlotinib dosing strategies for EGFR mutant Non-Small cell lung cancer patients. *PLoS One*, 10(11):e0141665.

[42] Loewe, S. (1953). The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung*, 3(6):285–290.

[43] Lu, J.-F., Eppler, S. M., Wolf, J., Hamilton, M., Rakhit, A., Bruno, R., and Lum, B. L. (2006). Clinical pharmacokinetics of erlotinib in patients with solid tumors and exposure-safety relationship in patients with non–small cell lung cancer. *Clinical Pharmacology & Therapeutics*, 80(2):136–145.

[44] Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W., Harris, P. L., Haserlat, S. M., Supko, J. G., Haluska, F. G., Louis, D. N., Christiani, D. C., Settleman, J., and Haber, D. A. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.*, 350(21):2129–2139.

[45] Mayer, I. A., Abramson, V. G., Formisano, L., Balko, J. M., Estrada, M. V., et. al. (2017). A Phase Ib Study of Alpelisib (BYL719), a PI3Kα-Specific Inhibitor, with Letrozole in $ER+/HER2-$ Metastatic Breast Cancer. *Clinical Cancer Research*, 23 (1): 26–34.

[46] Milton, D. T., Azzoli, C. G., Heelan, R. T., Venkatraman, E., Gomez, J. E., Kris, M. G., Krug, L. M., Pao, W., Rizvi, N. A., Dunne, M., and Others (2006). A phase I/II study of weekly high-dose erlotinib in previously treated patients with nonsmall cell lung cancer. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 107(5):1034–1041.

[47] Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*, 194(4260):23–28.

[48] Omuro, A. M. P., Kris, M. G., Miller, V. A., Franceschi, E., Shah, N., Milton, D. T., and Abrey, L. E. (2005). High incidence of disease recurrence in the brain and leptomeninges in patients with nonsmall cell lung carcinoma after response to gefitinib. *Cancer*, 103(11):2344–2348.

[49] Ouellet, D., Kassir, N., Chiu, J., Mouksassi, M.-S., Leonowens, C., Cox, D., DeMarini, D. J., Gardner, O., Crist, W., and Patel, K. (2016). Population pharmacokinetics and exposure–response of trametinib, a MEK inhibitor, in patients with BRAF V600 mutation-positive melanoma. *Cancer Chemother. Pharmacol.*, 77(4):807–817.

[50] Pao, W., Miller, V., Zakowski, M., Doherty, J., Politi, K., Sarkaria, I., Singh, B., Heelan, R., Rusch, V., Fulton, L., Mardis, E., Kupfer, D., Wilson, R., Kris, M., and Varmus, H. (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated

with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. U. S. A.*, 101(36):13306–13311.

[51] Ritz, C., Baty, F., Streibig, J. C., and Gerhard, D. (2015). Dose-Response analysis using R. *PLoS One*, 10(12):e0146021.

[52] Rugo, H. S., Andre, F., Rubovszky, G., Kaufman, B., Inoue, K., Takahashi, M., Shimizu, S., Ciruelos, E. M., Campone, M., Conte, P. F., Iwata, H., Loibl, S., Mayer, I. A., Juric, D., Longin, A.-S., Mills, D., Wilke, C., and Sellami, D. B. (2017). A phase 3 study of alpelisib (ALP) plus fulvestrant (FUL) in men and postmenopausal women with hormone receptor-positive (HR+), human epidermal growth factor receptor 2-negative (HER2-) ABC progressing on or after aromatase inhibitor (AI) therapy: SOLAR-1. *J. Clin. Orthod.*, 35(15_suppl):TPS1111–TPS1111.

[53] Togashi, Y., Masago, K., Fukudo, M., Terada, T., Fujita, S., Irisa, K., Sakamori, Y., Kim, Y. H., Mio, T., Inui, K.-I., and Mishima, M. (2010). Cerebrospinal fluid concentration of erlotinib and its active metabolite OSI-420 in patients with central nervous system metastases of non-small cell lung cancer. *J. Thorac. Oncol.*, 5(7):950–955.

[54] Vermes, I., Haanen, C., Steffens-Nakken, H., and Reutellingsperger, C. (1995). A novel assay for apoptosis flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled annexin V. *J. Immunol. Methods*, 184(1):39–51.

[55] Wong, K.-K., Fracasso, P. M., Bukowski, R. M., Lynch, T. J., Munster, P. N., Shapiro, G. I., Jänne, P. A., Eder, J. P., Naughton, M. J., Ellis, M. J., Jones, S. F., Mekhail, T., Zacharchuk, C., Vermette, J., Abbas, R., Quinn, S., Powell, C., and Burris, H. A. (2009). A phase I study with neratinib (HKI-272), an irreversible pan ErbB receptor tyrosine kinase inhibitor, in patients with solid tumors. *Clin. Cancer Res.*, 15(7):2552–2558.

[56] Yu, H. A., Sima, C., Feldman, D., Liu, L. L., Vaitheesvaran, B., Cross, J., Rudin, C. M., Kris, M. G., Pao, W., Michor, F., and Riely, G. J. (2017). Phase 1 study of twice weekly pulse dose and daily low-dose erlotinib as initial treatment for patients with EGFR-mutant lung cancers. *Ann. Oncol.*, 28(2):278–284.

[57] Zheng, H.-C. (2017). The molecular mechanisms of chemoresistance in cancers. *Oncotarget*, 8(35):59950–59964.

Supplementary Figures

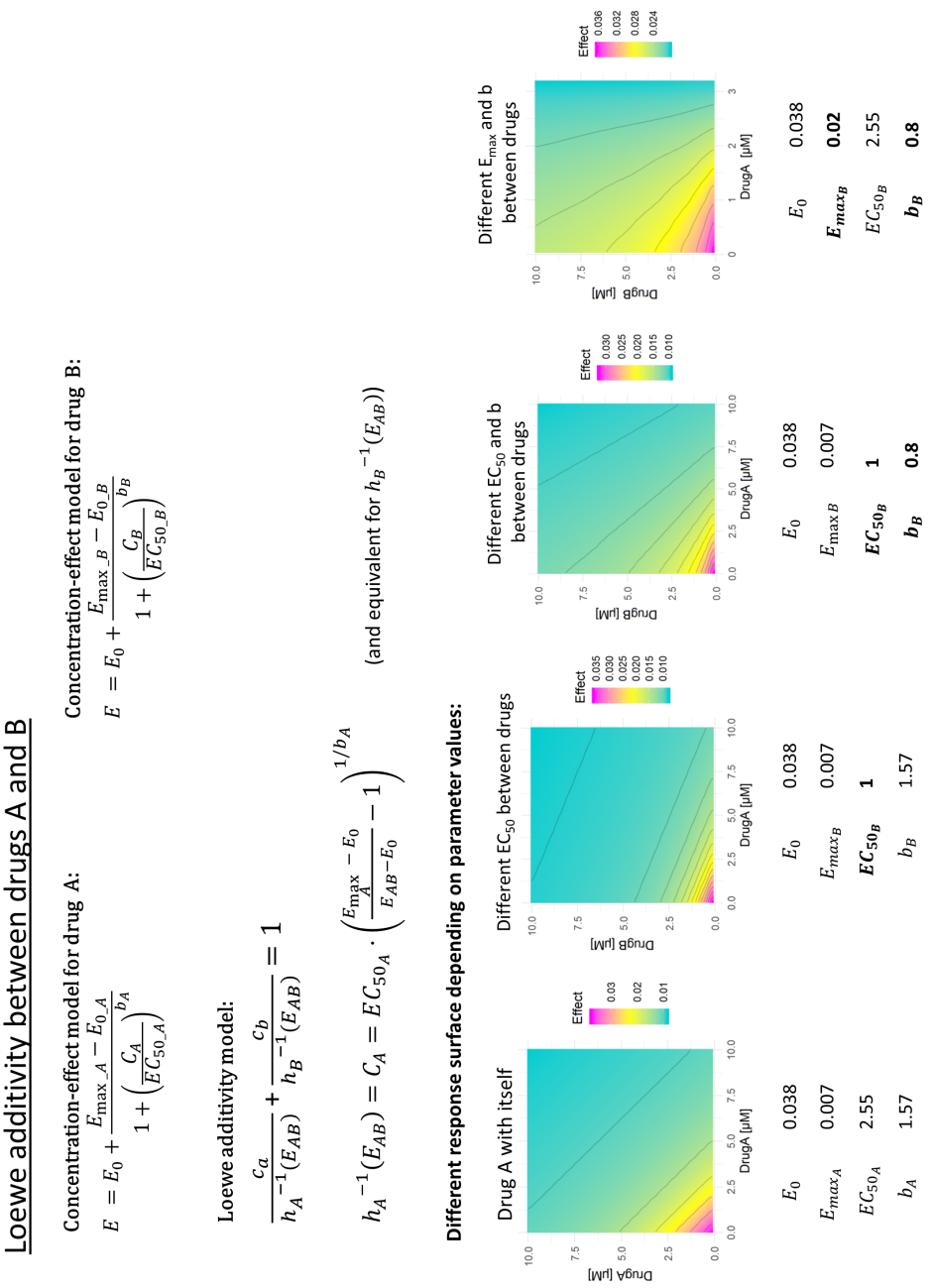
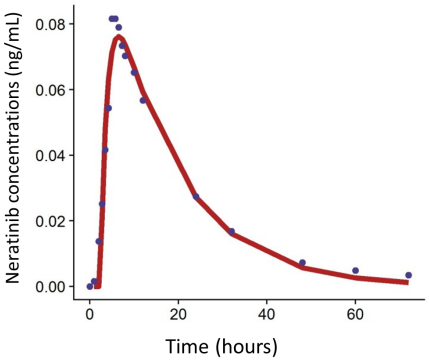


Fig. S3.1 Loewe additivity between drugs A and B. Loewe additive model response at any combined concentration is calculated from the sigmoidal fits of the single-agent response curve, which in this figure corresponds to a four parameter logistic equation. The model parameters used to simulate each response surface are indicated below each graph.



Parameter	Estimate
CL (L/h)	157.179
V (L)	2402.736
KA (h^{-1})	0.6029
ALAG (h)	2.3486
AIC	-118.66

Fig. S3.2 Estimation of the pharmacokinetic parameters driving the time course of neratinib concentration. The solid line represents the prediction of the model and the solid points represent the observations. CL indicates total drug clearance, V is the apparent volume of distribution, KA is the first-order absorption rate constant, ALAG is the lag time associated with the absorption of the drug and AIC is the Akaike Information Criterion.

Supplementary Material

Equations for the expected number of cells assuming cross-resistance to drugs

To model cross resistance to drugs, each of the resistant cell types are allowed to gain additional mutations and become resistant to more than one compound. A resistant type i cell can thus further mutate to form a type ia cell with probability u_{ia} (u_3 in Figure 1A). The first order birth and death rates of the ia cell are given by b_{ia} and d_{ia}. Thus, the time evolution equations of the expected number of cells in each clone are given by:

$$\begin{aligned}
\langle n_i(t)\rangle &= N_0 \cdot e^{-\int_0^t [b_0(s)\cdot(1-\sum_{z=1}^{N} u_{i_z})-d_0(s)]ds} && for\ i=0\\
&= \frac{N_i + \int_0^t [e^{-\int_0^s [b_i(\tau)(1-\sum_{cr} u_{ia_{cr}})-d_i(\tau)]d\tau} \cdot u_i \cdot b_0(s) \cdot N_0 \cdot e^{\int_0^t [b_0(\tau)\cdot(1-\sum_z u_{i_z})-d_0(\tau)]d\tau}]ds}{e^{-\int_0^t [b_i(s)(1-\sum_{cr} u_{ia_{cr}})-d_i(s)]ds}} && for\ i>0\\
\langle n_{ia}(t)\rangle &= \frac{N_{ia} + \int_0^t [e^{-\int_0^s [b_{ia}(\tau)-d_{ia}(\tau)]d\tau} \cdot u_{ia} \cdot b_i(s) \cdot n_i(s)]ds}{e^{-\int_0^t [b_{ia}(s)-d_{ia}(s)]ds}} && for\ i>0
\end{aligned}$$

Non-parametric models to estimate the effect of drug combination data

To estimate the regression surface of the in-vitro drug combination data, nonparametric fitting methods like Generalized Additive Models (GAMs) [5] or locally weighted scatterplot smoothing (loess) regression [1] were used. GAMs are a nonparametric extension of generalized linear models (GLMs) [6].

Here, a general nonparametric function (e.g. cubic splines) that relates the predicted drug effect values to the drug concentrations is defined [8]. Loess is a particular implementation of local polynomial smoothing which fits simple models to localized subsets of the data. As nonparametric methods, GAMs and loess regression are data-driven rather than model-driven; that is, they allow the data to determine the shape of the response curves. Thus, these models are used to describe the relation between drug concentrations and the growth and death rates of cancer cells without assuming the data must fit some distribution shape. Although these methods are very flexible, they are not biologically interpretable. Even so, they are a very powerful exploratory tool which often shows relatively complex relations between a dependent variable and more than one independent variable without being limited by the shapes available in a parametric model.

This approach allowed us to incorporate drug combination data into our evolutionary framework and explore the effect of multidrug dosing schedules in the evolution of cancer cells.

Assessing drug synergy/antagonism

In order to quantify the degree of synergy/antagonism between two compounds, the typical approach is to compare their measured combination effect to a null reference model of no interaction, i.e. the expected response assuming no interaction between the two compounds. If the combination response is greater than what is expected by the reference model, the combination is classified as synergistic, while antagonism is defined when the combination produces less than the expected effect.

There are several well-known conventional approaches that define different null models to assess drug synergy/antagonism. The Loewe Additivity model [42] is one of the most commonly used models to quantify a zero-interactive state for the combination of two drugs. This model is based on the assumption that a drug cannot interact with itself and defines synergy/antagonism as a combined inhibitory effect that is greater/lower than the sum of the individual effects of the drugs. The general equation of this model is:

$$\frac{da}{DA} + \frac{db}{DB} = 1 \tag{3.11}$$

where da and db are the dose (or concentrations) of drug A and B in the combination that produce an effect E_{AB} and DA and DB represent the single doses of drug A and B necessary to reach the same effect E_{AB}. Isobole analysis [7, 2] and the Greco model [4] are methods derived from this equation.

Loewe additive model response at any combined concentration is calculated from the sigmoidal fits of the single-agent response curves. In order to obtain the concentrations of each drug given as a single agent that elicits an effect E_{AB}, an inverse hill equation (h^{-1}) with parameters obtained from

the individual dose-response curves using a common baseline value (effect when there is no drug concentration) is employed:

$$\frac{da}{h_A^{-1}(E_{AB})} + \frac{db}{h_B^{-1}(E_{AB})} = 1 \tag{3.12}$$

$$h_A^{-1}(E_{AB}) = C_A = EC_{50_A}(\frac{E_{max_A} - E_0}{E_{AB} - E_0} - 1)^{1/b_A} \quad (and\, equivalent\, for\, h_B^{-1}(E_{AB})) \tag{3.13}$$

Classical Loewe additivity model assumes that the drugs in the combination have equal individual baseline and maximum effects. However, this method can be extended to account for different drug maximal responses [9]. We refer to this extended method as Generalized Loewe additivity model. In the Supplementary Figure S2 an isobole analysis done for the combined action of two hypothetical drugs with differing single-agent model parameters is shown.

Highest Single Agent (HSA), also known as Gaddum's non-interaction model [3], is another popular model which defines a "independent action" of the drugs when the predicted effect of a combination is that of the one most effective drug alone. The HSA zero interaction model predicts the combined effect EAB for two single agents with effects EA and EB as:

$$E_{AB} = max(E_A, E_B) \qquad (or\, min(E_A, E_B)\, if\, the\, monotherapy\, curves\, are\, decreasing) \tag{3.14}$$

where E_A and E_B are measured on the monotherapy dose–response curve of drug A and B respectively. According to this model, any combined effect stronger than the effect of a single drug is called 'synergism' and a weaker effect 'antagonism'.

References

[1] Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991). Local regression models. chapter 8 of statistical models in S (edited by JM chambers and TJ hastie), 309-376.

[2] Cokol, M., Chua, H. N., Tasan, M., Mutlu, B., Weinstein, Z. B., Suzuki, Y., Nergiz, M. E., Costanzo, M., Baryshnikova, A., Giaever, G., Nislow, C., Myers, C. L., Andrews, B. J., Boone, C., and Roth, F. P. (2011). Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.*, 7:544.

[3] Gaddum, J. H. (1985). *Gaddum's Pharmacology*. Oxford University Press, USA.

[4] Greco, W. R., Park, H. S., and Rustum, Y. M. (1990). Application of a new approach for the quantitation of drug synergism to the combination of cis-diamminedichloroplatinum and 1-β-D-arabinofuranosylcytosine. *Cancer Res.*

[5] Hastie, T. J. (2017). Generalized additive models. In *Statistical models in S*, pages 249–307. Routledge.

[6] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, Second Edition*. CRC Press.

[7] Tallarida, R. J. (2006). An overview of drug combination analysis with isobolograms. *J. Pharmacol. Exp. Ther.*, 319(1):1–7.

[8] Troconiz, I. F., Sheiner, L. B., and Verotta, D. (1994). Semiparametric models for antagonistic drug interactions. *J. Appl. Physiol.*, 76(5):2224–2233.

[9] Van der Borght, K., Tourny, A., Bagdziunas, R., Thas, O., Nazarov, M., Turner, H., Verbist, B., and Ceulemans, H. (2017). BIGL: Biochemically intuitive generalized loewe null model for prediction of the expected combined effect compatible with partial agonism and antagonism. *Sci. Rep.*, 7(1):17935.

4

Predicting circulating biomarker response and its impact on the survival of advanced melanoma patients treated with adjuvant therapy

Predicting circulating biomarker response and its impact on the survival of advanced melanoma patients treated with adjuvant therapy

Itziar Irurzun-Arana[1,2], Eduardo Asín Prieto[1,2], Salvador Martín-Algarra[2,3], Iñaki F. Trocóniz[1,2]

1 Pharmacometrics & Systems Pharmacology group, Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Navarra, Spain.

2 IdiSNA, Navarra Institute for Health Research, Pamplona, Navarra, Spain.

3 Department of Medical Oncology, Clínica Universidad de Navarra, Pamplona, Navarra, Spain.

Manuscript in preparation

Abstract

Advanced melanoma remains a disease with poor prognosis. Several serologic markers have been investigated to help monitoring and prognostication, but to date only lactate dehydrogenase (LDH) has been validated as a standard prognostic factor biomarker in malignant melanoma by the American Joint Committee on Cancer.

In this work, we have constructed a semi-mechanistic model to explore the relationship between the time course of several circulating biomarkers and overall or progression free survival in advanced melanoma patients treated with adjuvant high-dose interferon-$\alpha 2b$. Additionally, due to the adverse interferon tolerability, a semi-mechanistic model describing the side effects of the treatment in the absolute neutrophil counts was built in order to simultaneously analyze the benefits and toxic effects of this treatment.

The relative change from baseline of LDH was the most significant predictor of the overall survival of the patients. Unfortunately, there was no significant difference in the proportion of patients with elevated serum biomarkers between the patients who recurred and those who remained free of disease. Thus, no link was established between biomarker levels and progression-free survival.

We believe that the modeling framework presented in this work of circulating biomarkers and adverse effects could constitute an additional strategy for disease monitoring in advance melanoma patients.

4.1 Introduction

According to the American Cancer Society, the incidence rate of melanoma has been rising for the last 30 years. Although the disease accounts for only about 1% of skin cancers, it is responsible for the death of the vast majority of these patients making it the most aggressive neoplasm of the skin [8].

Since 1995, immunotherapy based on interferon-α containing regimens has been used as an adjuvant therapy to surgery for patients diagnosed of American Joint Committee on Cancer (AJCC) stage IIB, IIC or III melanoma after the Eastern Cooperative Oncology Group (ECOG) 1684 trial showed that a high-dose regimen of Interferon α-2b (IFN-$\alpha 2b$) led to a significant prolongation of progression-free survival and overall survival (PFS and OS, respectively) compared to the control group [13]. Although new therapeutic strategies are emerging for advanced melanoma in recent years thanks to the FDA approval of several new immunotherapy and targeted drugs, treatment with IFN-$\alpha 2b$ still constitutes one of the alternatives in the therapeutic arsenal in many hospitals and health care centers. However, due to the toxicity and the evidence that only a subgroup of patients can benefit from this treatment, acceptance of IFN-$\alpha 2b$ among physicians is limited.

In order to adequately treat melanoma patients, it is important to study those factors related to the prognosis and outcome of the disease. As reflected in recent studies, the most important prognostic factors that could predict the outcome of melanoma patients include the vertical tumor thickness known as Breslow's index, the presence of ulceration, the mitotic rate, the location of distant metastases, as well as the levels of serum lactate dehydrogenase (LDH) [8]. Other serum biomarker levels that have been proposed as possible prognostic factors are the melanoma-inhibiting activity (MIA) and the calcium binding protein S100B [17], but no consensus exists on their prognostic capability.

Proper assessment of the predictive capacity of biomarkers longitudinal data should be done in the context of mechanistic computational models linking them with clinical outcome. Biomarker trajectories are usually not linear and show great variability across individuals. Consequently, a non-linear mixed effects (NLME) modeling approach provides a valuable option to handle and model this type of dynamic behavior. In NLME models, individual profiles are characterized by a common structural model with fixed population parameters and a statistical model with random effects to allow the parameters to vary within the patient population. In this work, longitudinal biomarker data has been described based on semi-mechanistic pharmacokinetic-pharmacodynamic (PKPD) type models and linked to the PFS and OS. Recent efforts have shown that this approach is feasible to identify robust markers that allow the selection of patients that could obtain a therapeutic benefit from the different anticancer treatments and to improve the prediction of their survival [4, 20, 6].

Therefore, in this study we aim to establish a quantitative treatment-biomarker-survival modeling framework using nonlinear mixed effects PKPD modeling to link the survival of advanced melanoma

patients with LDH, MIA and/or S100B protein kinetics following IFN-$\alpha 2b$ administration. In addition and taking into account the toxicity associated to IFN-$\alpha 2b$ administration, neutropenic effects were also described mechanistically [7] in the current evaluation providing a highly valuable approach in which to evaluate possible predictors of clinical response while minimizing adverse effects.

4.2 Material and Methods

4.2.1 Patient characteristics and data collection

In this retrospective study, data related to different biomarker levels and patient survival were obtained from the medical records of 48 patients diagnosed with advanced melanoma and treated in the University Clinic of Navarra (Pamplona, Spain).

Adult patients with histologically documented AJCC stage IIB, IIC, or III primary cutaneous melanoma were included in the dataset. All the patients were treated with adjuvant high-dose IFN-$\alpha 2b$ between 2004 and 2013. The high-dose regimen followed the Kirkwood scheme [13]: intravenous administration of 20 MU/m2/day at the induction phase (5 days/week during 4 weeks) followed by subcutaneous injections of 10 MU/m2/day during the maintenance phase (3 days/week during 48 weeks). Blood samples for drug quantification and tumor assessment measurements during treatment were not available.

Table 4.1 summarizes physiopathological and demographic characteristics of the patients included in the study and Table 4.2 summarizes the main adverse events reported during IFN-$\alpha 2b$ therapy.

Blood samples for measurement of LDH, MIA and protein S100B were collected from each patient before, meanwhile and after therapy. For MIA and S100B levels, observations corresponding to 9 and 10 patients of the database were not reported, respectively. A total of 954/383/405 LDH/MIA/S100B observations were included in the analysis, where each patient contributed a mean of 19/10/10 samples (range 1-57/1-31/1-33).

4.2.2 Data analysis

A population joint sequential modeling approach was used for the development of the treatment-biomarker-survival framework [5]. First, the relationship between treatment and biomarkers dynamics was characterized, and then their predicted time profiles were used to characterize the hazard rates and subsequently PFS and OS. For the continuous (biomarker levels, and absolute neutrophils counts) and non-continuous (PFS, OS) response data the first-order conditional estimation method with interaction and the Laplacian estimation method were used, respectively, for parameter estimation in NONMEM 7.3 (Icon Development Solutions, Ellicott City, MD).

The continuous data of the different biomarkers and the absolute neutrophil counts (ANC) were logarithmically transformed for the analysis. The developed models share a common architecture constituted by a structural model and a statistical component where (i) between-subject variability (BSV) was modeled exponentially and (ii) residual variability was described using a proportional or

Table 4.1 Demographic characteristics and diagnostic values of the patients*

Demographic Characteristics	***Overall population (N = 48)***
Gender (M/F)	25/23
Age at melanoma diagnosis (years)	50 [21 – 74]
Body weight at first IFN dose (kg)	73 [45 – 108]
Height (cm)	168.5 [148 – 188]
BSA (m2)	1.825 [1.37 – 2.31]
Diagnosis values	
Location of primary lesion	
Face	3
Trunk	15
Extremity (Upper/lower)	4/19
Other	4
NR	3
Type of melanoma	
Amelanotic melanoma	1
Superficial spreading melanoma	16
Acral lentiginous melanoma	2
Maligna melanoma	14
Nodular melanoma	15
Laterality (Right/Left/NR)	11/14/23
Local recurrence (Yes/No/NR)	05/02/41
Diagnostic - Pathological stage AJCC 2009b	
IB	6
IIA/IIB	1/6
III/IIIA/IIIB/IIIC	2/8/9/4
NR	12
First dose - Pathological stage AJCC 2009b	
IIB	4
III/IIIA/IIIB/IIIC	10/8/14/7
NR	5
SLNB	
Yes (Positive cases)	32 (21)
No	16
History of complete lymphadenectomy	
Yes (Positive cases)	44 (41)
No	3
NR	1
Breslow thickness (mm)	
< 1	5
≥ 1 to < 2	10
≥ 2 to < 4	13
≥ 4	12
Clark level	
II	2
III	12
IV	22
V	2
NR	10
Ulceration (Yes/No/NR)	8/22/18
Extracapsular extension (Yes/No/NR)	3/35/10
Satellite lesions (Yes/No/NR)	3/24/21
BRAF Mutation (Yes/No/NR)	08/08/32
ECOG performance status (before therapy)	
0	14
1	21
NR	13

M: male; F: female; BSA, Body Surface area; AJCC, American Joint Committee on Cancer; SLNB, Sentinel Lymph Node Biopsy; ECOG, Eastern Cooperative Oncology Group; NR: Not reported.
*Continuous variables are expressed as median [range] whereas categorical variables are expressed as number of cases.

Table 4.2 Main adverse events reported during IFN therapy.

Main adverse events*	***Induction phase***	***Maintenance phase***
Neutropenia	13	6
Thrombocytopenia	2	1
Increased transaminases	12	3
Hepatotoxicity	8	1
Fatigue	9	8
Osteoarticular pain	1	2
Influenza-like symptoms	2	3
Fever	3	1
Headache	4	-
Anorexia	1	5
Depression	3	4
Nausea	2	2

*Other adverse events reported were: dermal events (cellulitis, dermatitis, skin dryness and alopecia), neurological events (anxiety, somnolence, insomnia, dizziness, recurrent syncope), weight loss and hyperthyroidism.

an additive error model on the log-transformed data corresponding to the biomarker and ANC levels respectively.

Model selection

Model selection during model building included comparison of the objective function value which is approximately equal to minus twice the log(likelihood) (-2LL) and inspection of graphical diagnostics. For application of the -2LL ratio test in the case of comparing nested models, a significance level of $P<0.01$ was used, corresponding to a decrease in -2LL of at least 6.63 when one extra parameter was added. Non-nested models were compared using the Akaike information criteria (AIC) [1].

Model evaluation

Evaluation was performed through simulation-based diagnostics by performing visual predictive checks (VPC) [3]. VPCs evaluate the model's ability to describe the median tendency and variability in the observed data. To this end, the original dataset was simulated 1000 times by sampling new sets of individual parameters from the estimated population parameter distributions. Then, 95% prediction intervals were derived from the simulation results, and compared with the 5th, 50th and 95th percentiles of the observed data. The results of the VPCs can also be normalized by the typical population prediction, creating the so-called prediction-corrected VPCs.

Precision of parameter estimates was obtained from the analysis of 500 bootstrap datasets. Briefly, in a bootstrap analysis, the original dataset is replaced to produce another dataset of the same size but with a different combination of individuals. This re-sampled database is then used to re-estimate

the population and variability parameters of the model. Lastly, median values and 95% confidence intervals of the re-estimated parameter distribution are calculated.

Software and tools

Preprocessing of the data, additional simulation exercises and graphical and other statistical analyses, including predictive checks and bootstrap analyses, were performed with Perl-speaks-NONMEM (PsN) software [15], Simulx (http://simulx.webpopix.org/), R version 3.4.3 (http://www.R-project.org/) and Rstudio version 1.1.456 (http://www.rtudio.com/).

Model for biomarker response

As above mentioned, NLME models were used to characterize the longitudinal LDH, MIA and S100B protein concentrations over time.

As no PK data of interferon therapy were available from the patients, a K-PD modeling approach [10] was used to study the link between the interferon dosing rate and the biomarker dynamics. A preliminary exploratory analysis showed that the decrease in biomarker levels occurred with some delay after treatment administration (see Figure 4.2A) that was handled incorporating a series of transit compartments. Transit of the pharmacodynamic signal elicited by interferon through the chain of compartments was characterized by the first order rate constant k_{tr} defined as $(n+1)/MTT$, where n is the number of transit compartments and MTT the mean transit time between compartments.

In the absence of treatment, an exponential tumor growth governed by a first-order proliferation rate constant (k_{prol}) was defined in the following form:

$$\frac{dTA}{dt} = k_{prol} \cdot TA \tag{4.1}$$

where TA (Tumor Activity) represents the unobserved tumor progression dynamics. IFN-$\alpha 2b$ therapy induced tumor shrinkage, and hence, the final equation for TA was expressed as a balance between tumor growth and drug-induced tumor death:

$$\frac{dTA}{dt} = k_{prol} \cdot TA - f_{drug} \cdot TA \tag{4.2}$$

Different models for drug effects (f_{drug}) were explored including linear, E_{max} and sigmoidal models. The value of TA at diagnosis (TA_0) was arbitrarily set to 1.

Lastly a turn-over model assuming that the circulating levels of biomarkers are a function of (i) a synthesis process governed by TA and the first-order rate constant k_{in}, and (ii) an elimination process controlled by the first-order rate constant k_{out} as shown in the expression below:

$$\frac{dBiomarker_j}{dt} = k_{in_j} \cdot TA - k_{out_j} \cdot Biomarker_j \tag{4.3}$$

where j represents each of the biomarkers (LDH, MIA and S100B). The initial condition for biomarker values was estimated as due to tumor progression steady-state condition did not hold.

Each biomarker's longitudinal data were fitted separately using the model equations described above. Afterwards, the kinetics of the three biomarkers was combined in the same analysis to evaluate their contribution in the clinical outcome of the patients.

Models for progression-free survival and overall survival

Kaplan-Meier curves were generated to perform a first exploratory analysis of the overall and progression-free survival stratified based on well-established clinical variables whose differences across patients were tested for significance by the log-rank test (Figure 4.1).

PFS and OS were modeled as time to event response data using parametric survival analyses. Time frame was considered between diagnosis and (i) time at which the patient showed disease progression or died and (ii) last recorded time (right censored). Interval censored for the case of PFS response was not considered in this evaluation.

Different distributions (exponential, Weibull and Gompertz) were used to describe the hazard rate, $hz(t)$, which is defined as the instantaneous risk of dying/recurring at each time provided that the patient lives/is free of disease to that time, with the sole restriction of being no-negative. In contrast to the hazard function, the survival function indicates the probability that the event of interest has not yet occurred by time t (the patient is still alive or free of disease) and therefore if the hazard function is known, the survival probability is automatically determined as follows:

$$S(t) = exp(-\int_0^t hz(t)) \tag{4.4}$$

where $-\int_0^t hz(t)$ represents the cumulative hazard.

Time-varying covariates, as the predicted time course of the biomarkers, were included in the model as modulators of $hz(t)$. Parameters describing $hz(t)$ has no associated BSV as each patient contributed with a single measurement.

The effect of the predicted dynamics for the three biomarkers on $hz(t)$ were tested alone or in combination to explore whether their absolute or relative change from baseline over time were predictive of OS/PFS. For the estimation of the parameters linking the survival and the biomarker

model, the population parameters from the previously selected biomarker model were fixed and the corresponding observed levels were retained together with the PFS and OS data (PPP&D method [21]).

Model for neutropenic adverse effects

A semi-mechanistic model for myelosupression [7] was used to characterize the dynamics of the absolute neutrophil circulating counts under IFN-$\alpha 2b$ therapy. Briefly, in this model neutrophil development is determined by different physiological processes: (i) a self-renew first-order process of the precursor cells (ii) a maturation chain comprising three transit compartments (iii) a homeostatic regulation that modulates the proliferation of the precursor cells as a function of the change of ANC relative to the value at baseline (ANC_0), and finally (iv) a first-order elimination of ANC. As said before, no interferon PK data were available, and therefore, a K-PD model [10] was used to link the dosing rate to drug effects.

The model structure is defined by the following set of ordinary differential equations:

$$\frac{dIFNa2b}{dt} = -K_e \cdot IFNa2b \tag{4.5}$$

$$\frac{dProl}{dt} = k_{PROL} \cdot Prol \cdot (1 - E_{DRUG}) \cdot \left(\frac{ANC_0}{ANC}\right)^{\gamma} - k_{TR} \cdot Prol \tag{4.6}$$

$$\frac{dTransit1}{dt} = k_{TR} \cdot Prol - k_{TR} \cdot Transit1 \tag{4.7}$$

$$\frac{dTransit2}{dt} = k_{TR} \cdot Transit1 - k_{TR} \cdot Transit2 \tag{4.8}$$

$$\frac{dTransit3}{dt} = k_{TR} \cdot Transit2 - k_{TR} \cdot Transit3 \tag{4.9}$$

$$\frac{dANC}{dt} = k_{TR} \cdot Transit3 - k_{circ} \cdot ANC \tag{4.10}$$

where K_e represents the first-order elimination rate constant of interferon after administration, k_{PROL} is the first-order rate of proliferation on precursor cells (Prol), k_{TR} is the first order rate constant governing the transit of immature neutrophils between transit compartments, k_{circ}, is the first-order rate constant of elimination of ANC and γ is the parameter modulating the feedback mechanism. The transit rate was defined as $k_{TR} = (n+1)/MTT_{ANC}$ where MTT_{ANC} is the mean maturation time and n is the number of transit compartments, which was three in this model. As no information was gathered from the precursor and immature cells, it was assumed that, at baseline, their number of cells were equal to ANC_0, and therefore the parameter values for k_{PROL}, k_{TR} and k_{circ} were defined to be equal. Both a linear and a sigmoidal E_{max} function of the predicted levels of IFN-$\alpha 2b$ were evaluated for drug effects (E_{DRUG}), which were assumed to act by reducing the proliferation rate of the neutrophils.

In order to reduce the number of parameters to estimate and improve model stability, the parameters reported in the original work [7] for MTT and γ were used, as the authors demonstrated that the estimates of the system related parameters showed consistency across different anti-cancer agents. The final parameters to be estimated were reduced to ANC_0, parameters measuring drug effects and those quantifying random effects.

Covariate selection

Covariate model selection was performed using the Stepwise Covariate Model-building (SCM) tool in PsN [14], which consists on a forward covariate inclusion followed by a backward deletion approach. Specifically, this technique consist on creating a full model by combining the covariates identified as significant (p<0.05) and once the full model is established, each potential covariate is individually removed to see if the value of -2LL significantly increases (p<0.01).

Patient characteristics are listed in Table 4.1. For those covariates that were correlated between them, as it was the case for weight, height and body surface area (BSA), only the most relevant covariate with regard to usual dose adjustments in the clinic, in this case BSA, was included in the analysis. Therefore, the following patient's characteristics measured at baseline were explored for inclusion in the model (the covariates were tested in all the model parameters): Breslow thickness, presence of ulceration (yes vs. no), age, body surface area, type of melanoma (horizontal growth phase vs vertical growth phase) and ECOG performance status. Other a priori important clinical covariates like the presence of BRAF mutation or the mitotic rate were not studied as the number of missing data was high. The categorical level of invasion known as the Clark index was neither included in the analysis as almost every patient had reported a level of IV.

Covariates were tested for significance following the general model:

$$TVP = \theta_n \cdot \prod_{1}^{m} g(cov_m, cov_{m,ref}, \theta_m) \cdot \prod_{1}^{p} (1 + \sum_{cat=2}^{ctg} \theta_{p,cat}) \tag{4.11}$$

where the typical value of a parameter (TVP) was described as a function of m continuous (cov_m) and p categorical covariates (cat) with a total number categories of ctg. θ_n describes the n^{th} typical parameter value for an individual with covariate values equal to the reference values: [$(cov_m = cov_{m,ref})$ and cat = 1] where $cov_{m,ref}$ refers to the median value across the studied population. g refers to the different linear and non-linear functions explored for the relationship between the values of cov_m and $cov_{m,ref}$, and θ_m and $\theta_{p,cat}$ are parameters quantifying the magnitude of the covariate-parameter relationship.

4.3 Results

A total of 30 (62%) and 21 (43%) patients completed the induction and the maintenance phase, respectively. In total, 17 patients (35%) had at least one dose reduction during the induction or maintenance phase due to adverse events and 25 patients (52%) had dose delays for the same reason, demonstrating the high toxicity of IFN-$\alpha 2b$ therapy.

4.3.1 Exploratory analysis

The median overall survival of the patients in the dataset was 270 weeks. A first exploratory analysis of the dataset showed that patients with high LDH, MIA and S100B levels have the poorest outcomes as indicated by the Kaplan–Meier curves of OS shown in the top of Figure 4.1 and Supplementary Figure S4.1. However, the Kaplan-Meier analysis and log-rank tests corresponding to the PFS response did not show significant results (p>0.01) when stratifying by high and low biomarker values at the time of disease progression (Figure 4.1 and Supplementary Figure S4.1 bottom). Other in principle relevant clinical covariates like the Breslow thickness, presence of ulceration, tumor extension (distal, localized or regional) or the value of the biomarkers before treatment initiation also showed no significant differences in OS or PFS (p>0.01) (Supplementary Figure S4.2 and S4.3). These findings suggest that a link might exist between biomarker dynamics during and after IFN-$\alpha 2b$ treatment and OS and not for PFS.

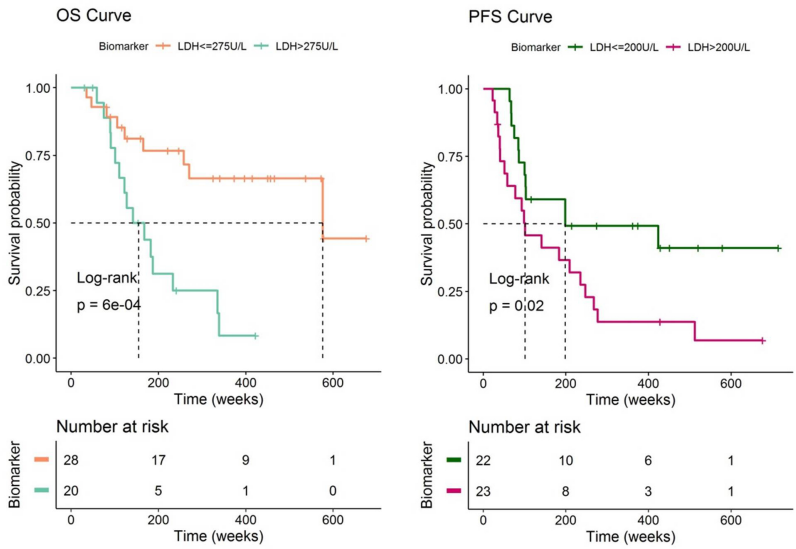


Fig. 4.1 Evaluation of the overall survival (OS) and progression-free survival (PFS) of the patients with high and low biomarker concentrations at the end of the study. MIA and S100 biomarker Kaplan Meier curves showed equivalent results (see supplementary figures).

The raw values for each biomarker are shown in Figure 4.2A, where the time course for one individual data and its treatment period (induction phase followed by the maintenance phase) has been highlighted. When looking at the whole range of observations, it is difficult to observe a general trend in the data. However, when the biomarker profiles are observed individually, a response to the therapy followed by a relapse after the treatment period can be detected. In this work, we intended to describe this trend and its link to the OS and PFS data using semi-mechanistic computational models.

4.3.2 Biomarker dynamics

Figure 4.2B provides a schematic representation of the model which is described by the following set of ordinary differential and algebraic equations:

$$\frac{dIFNa2b}{dt} = -k_{tr} \cdot IFNa2b \tag{4.12}$$

$$\frac{dTR}{dt} = k_{tr} \cdot IFNa2b - k_{tr} \cdot TR \tag{4.13}$$

$$\frac{dTA}{dt} = k_{prol} \cdot TA - f_{drug} \cdot TA \tag{4.14}$$

$$f_{drug} = k_{kill\_max} \cdot \frac{TR}{TR + TR_{50}} \tag{4.15}$$

$$\frac{dBiomarker_j}{dt} = k_{in_j} \cdot TA - k_{out_j} \cdot Biomarker_j \tag{4.16}$$

Drug effects were described with an E_{max} model were TR_{50} is the predicted pharmacodynamic signal generated by the treatment in the transit compartment eliciting half of maximum effect (k_{kill_max}). The rest of parameter abbreviations have been defined in Material and Methods.

Parameter estimates and their corresponding BSV are summarized in Table 4.3 . For the sake of parameter identifiability, the value of TR_{50} was fixed in the model of LDH and MIA dynamics after performing a sensitivity analysis study (data not shown). For S100B tumor marker the population estimate and BSV of MTT were also fixed to the values obtained in the model for LDH concentrations. None of the studied covariates had a significant effect on the model parameters.

The analysis of the Individual Weighted Residuals (IWRES) vs. time or predicted biomarker is shown in Supplementary Figure S4.4. Additionally, the three individual fits of a representative patient for each biomarker in Figure 4.2 and the results of the VPCs represented in Figure 4.3A demonstrated good agreement between observed and simulated data (only the VPC for LDH is shown).

With respect to parameter precision, none of the 95% confidence intervals for the model parameters reported in Table 4.3 (computed from the bootstrap analysis) included the value of zero, indicating

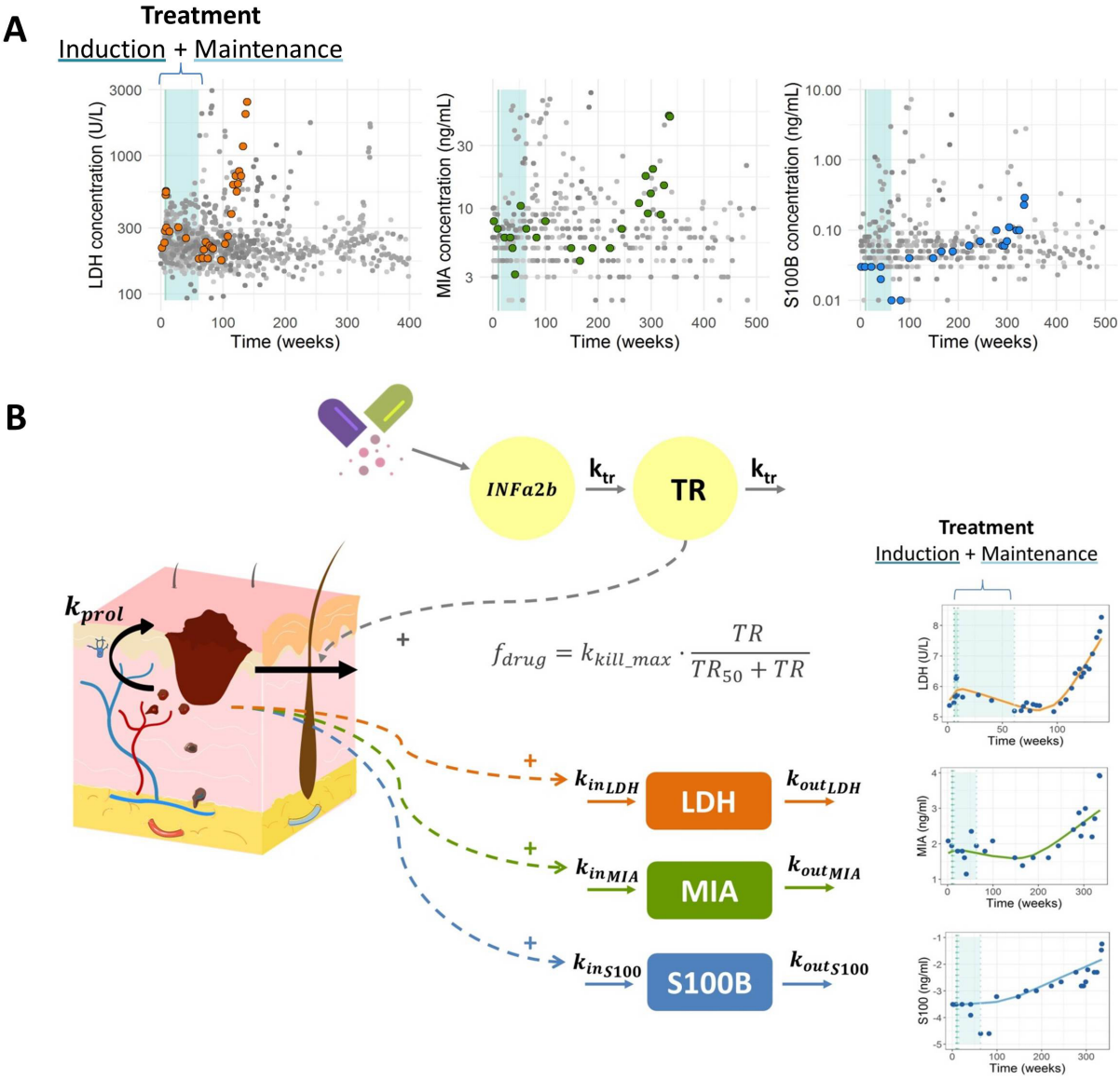


Fig. 4.2 A) Raw values (solid circles) of the different biomarker levels over time where the profile of one individual has been highlighted in color. The treatment period (induction phase followed by the maintenance phase) is shadowed in blue. B) Schematic representation of the K-PD model proposed for the IFN-$\alpha 2b$ effect on LDH, MIA and S100 levels (left) and three individual biomarker profiles (right) where solid circles represent biomarker observation values and solid lines indicate the prediction of the model. Parameter abbreviations: k_{prol}, first-order tumor proliferation rate; MTT, mean transit time; A_{50}, amount of drug producing 50% of the maximum elimination; k_{kill_max}, first-order tumor elimination rate; k_{in} first-order biomarker synthesis rate constant; k_{out}, first-order degradation rate constant.

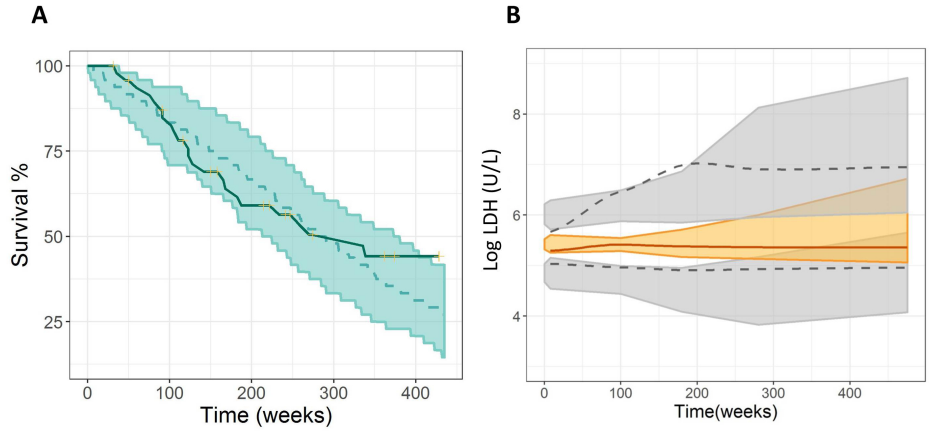


Fig. 4.3 Model evaluation: Visual Predictive Checks. A) Kaplan Meier plot of OS probability. The solid blue line represents raw data while the blue shaded area cover the 95% prediction interval calculated from 1000 simulated studies. B) VPC of the selected biomarker model. Median (solid line), 5^{th} and 95^{th} percentiles (dashed lines) of the observed data. 95% confidence Intervals for median (shaded colored area), 5^{th} and 95^{th} percentiles (shaded grey areas) of the simulated data.

that the data supported the degree of complexity of the final model selected. In all the models, k_{prol} and MTT showed a high BSV value and a wide range for the confidence interval of the BSV.

4.3.3 Survival model

Predicted biomarker dynamics over time were linked to the probability of survival as an argument of the baseline hazard function, which was best described using an exponential model with constant λ in the case of OS and with a Gompertz function in the case of PFS (Supplementary Figure S4.5). Relative change from baseline of LDH ($\Delta LDHrel$) was the most significant predictor of OS (p<0.001), however none of the biomarker dynamics significantly improved PFS predictions as previously suggested by the Kaplan-Meier curves from Figure 4.1 and Supplementary Figure S4.1. Additionally, none of the studied covariates (see Material and Methods for the information about the covariates tested in the model) influenced survival according to the univariate analysis done in PsN using the SCM tool and therefore none of them were included in the joint model afterwards.

The final survival model for has the following form:

$$hz(t) = \lambda \cdot e^{\beta \cdot \Delta LDHrel(t)} \tag{4.17}$$

where the term $e^{\beta \Delta LDHrel(t)}$ describes the change in hz elicited by the relative change from baseline of LDH for each individual i multiplied by the link parameter β. The estimated values for α and β are summarized in Table 4.3 and the corresponding VPC for OS is shown in Figure 4.3B. We only considered time up to 450 weeks after diagnosis to evaluate model performance through VPCs as for

Table 4.3 Final model parameter estimates

Model for biomarker response		
	LDH	
	Typical estimate	BSV: CV%
MTT (weeks)	22.3 (12.49-37.318)	59.8 (44.7-113.5)
$TR_{50}(U \cdot 10^6)$	23.7 (-)	-
LDH baseline (U/L)	225 (167.082-261.393)	31.3 (21.45-51.28)
$k_{prol}(weeks^{-1})$	0.0029 (0.00196-0.0053)	58 (51.75-224.27)
$k_{kill\_max}(weeks^{-1})$	0.0077 (0.0056-0.012)	34 (31.3-128.45)
$k_{out}(weeks^{-1})$	0.321 (0.189-0.822)	-
Residual error (%)	0.0521 (0.0376-0.0576)	NA*
	MIA	
MTT (weeks)	33.1 (28.04-36.9)	63.8 (41.6-83)
$TR_{50}(U \cdot 10^6)$	25.2 (-)	-
MIA baseline (ng/mL)	7.53 (5.57-7.85)	45.3 (27-54.8)
$k_{prol}(weeks^{-1})$	0.0028 (0.0022-0.0042)	75.6 (69.2-141)
$k_{kill\_max}(weeks^{-1})$	0.0058 (0.003-0.0061)	34.6 (25.5-62.4)
$k_{out}(weeks^{-1})$	0.369 (0.288-0.486)	-
Residual error (%)	0.248 (0.192-0.296)	NA
	S100B	
MTT (weeks)	22.3 (-)	59.8 (-)
$TR_{50}(U \cdot 10^6)$	19.9 (16.42-23.64)	-
S100B baseline (ng/mL)	0.0503 (0.038-0.0656)	42.4 (36.74-55.95)
$k_{prol}(weeks^{-1})$	0.0023 (0.0017-0.0025)	56.2 (45.8-73.48)
$k_{kill\_max}(weeks^{-1})$	0.0065 (0.005-0.0072))	-
$k_{out}(weeks^{-1})$	1.99 (1.71-2.49)	-
Residual error (%)	0.348 (0.3-0.424)	NA
OS model		
λ	0.00181 (0.0016-0.0026)	-
$\beta_{\Delta LDHrel}$	1.1 (0.59-1.9)	-
Myelosuppresion model		
$MTT_{ANC}(weeks)$	0.52	-
$ANC_0(10^9/L)$	3.41 (3.134-3.737)	32.4 (26.3-40)
$Slope(U \cdot 10^{-4})$	0.0425 (0.0392-0.055)	-
$K_e(weeks^{-1})$	0.389 (0.344-0.566)	52.4 (24.36-70.26)
γ	0.161	-
Residual error $(10^9/L)$	0.485 (0.465-0.501)	-

*NA: Not Applicable

90% confidence intervals (in parenthesis) were obtained from 500 bootstrap analyses. Estimates of between-subject variability (BSV) are shown as coefficients of variation. Parameter names are defined in the text.

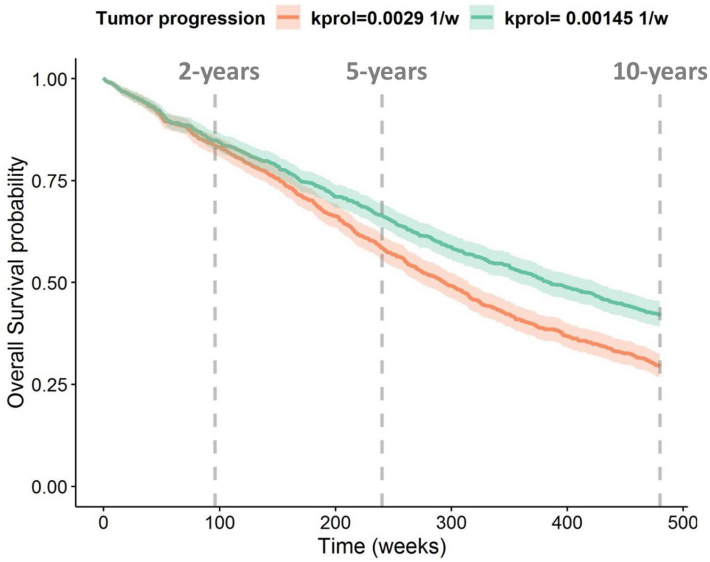


Fig. 4.4 Kaplan Meier overall survival stratified by different values of the tumor proliferation rate constant parameter (k_{prol}). The plot is the result of the simulation of 1000 individuals receiving the treatment in the same time period.

longer times only 7 individuals were remaining for a period of approximately 225 weeks more.

The predicted median 2-year and 5-year overall survival probability computed was 84.37% and 58.33% respectively, which were very similar to the observed values of 82.4% and 56.39% obtained from the 48 patients in our dataset. Additional simulation exercises where the therapy was administered in the same time period to all the individuals showed that a 50% decrease in tumor proliferation practically did not affect the 2-year survival rate, but increased the 5-year and 10-year rate a 13.7% and 42% respectively (see Figure 4.4).

4.3.4 Model for neutropenic adverse effects

Table 4.2 summarizes the main adverse events reported during interferon therapy. Due to the fact that neutropenia was one of the most reported and potential life threatening toxic effects, we decided to characterize this adverse response using the semi-mechanistic model from [7]. That semi-mechanistic myelosuppression model adequately described the time course of the log-transformed absolute neutrophil counts as illustrated by the prediction-corrected VPC from Figure 4.5A. The linear drug effect model showed significantly better fitting results compared with the E_{max} model ($p<0.01$). The final model included BSV in the ANC baseline parameter (ANC_0) and in the elimination rate constant (K_e) of the K-PD model (see Table 4.3 for parameter values). None of the studied covariates had a significant effect on the model parameters.

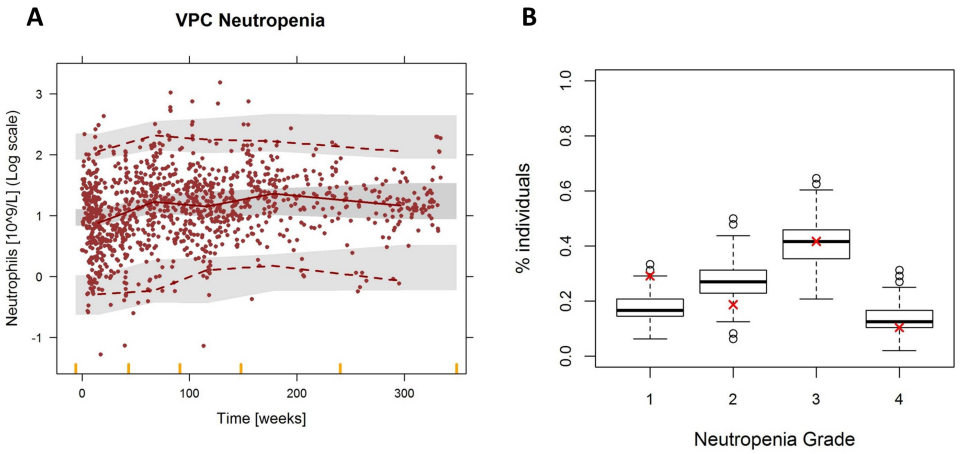


Fig. 4.5 Evaluation of the myelosuppression model for the absolute neutrophil counts (ANC) of the patients. A) Prediction-corrected visual predictive check. Solid circles represent observed ANC, solid lines represent the median of the observed data, and dashed lines the 2.5 and 97.5 percentiles of the observations. Shaded areas are the 95% confidence intervals based on the simulated data (n=1000) for the corresponding percentiles. B) Percentage of patient in grade 1,2,3 and 4 neutropenia (grade 1: >1.5 ANC, grade 2: 1-1.5 ANC, grade 3: 0.5-1 ANC, grade 4: <0.5 ANC). Boxplots summarize the result of the 500 simulations and the red cross represents the real percentage values from the dataset.

In Figure 4.5B the percentage of patients with grade 1, 2, 3 and 4 neutropenia calculated from five hundred simulated ANC vs time profiles were compared to the corresponding percentages derived from the observations. Results show that the model captures well severe grades of neutropenia.

4.4 Discussion

A joint model for the dynamics of circulating biomarkers and overall survival has been established and evaluated in patients with melanoma during treatment with IFN-$\alpha 2b$. Additionally, a myelosuppression model was also developed to evaluate the adverse effects of the IFN-$\alpha 2b$ therapy in the same cohort of patients. This framework enables to convert the individual biomarker levels into personalized predictions of survival while taking toxicity into account. All of the investigated biomarkers were significantly related to OS when evaluated one by one, but the relative change from baseline of LDH was identified as the most predictive of OS regarding objective function values. Although other studies also showed a significant association between LDH and PFS in melanoma [9], in our analysis none of the tumor marker dynamics significantly improved PFS predictions. Moreover, treatment with Interferon is more associated with an improvement in PFS rather than OS but our data did not allow us to characterize this link.

Serum LDH, which is a standardized biomarker routinely monitored in clinic, is also used to categorize patients with stage IV melanoma, as increased LDH values are known to be correlated with a poor outcome of the patients. However, the link between biomarker values and survival needs to be quantitatively characterized in order to allow for more meaningful predictions of patient prognosis. In this work, we add insights in this context by providing a treatment-biomarker-survival-toxicity framework where the effectiveness of alternate dosing regimens could be tested based on $\Delta LDHrel$ values and neutropenia. Figure 4.6 conceptualizes the computational framework as it shows the individual LDH and ANC profiles and the time course of the hazard rate differentiating by an individual who is alive at the last follow-up (patient 36) and an individual who died (patient 26). In this figure it can also be appreciated that the effect of the therapy on the ANC was much faster than the decrease in LDH. This justifies the differences found between the estimates for the k_{tr} parameter which had a value of 0.039 weeks^{-1} for the case of LDH and the K_e of the myelosuppression K-PD model which had a value of 0.389 weeks^{-1} (almost 10 times higher).

The treatment of advanced stage melanoma has evolved immensely in recent years with the success of new immunotherapies and targeted drugs [2]. Nowadays, it is well known that approximately 60% of melanomas harbor a mutation in the gene encoding for the serine/threonine protein kinase BRAF, which leaded to the development of selective BRAF inhibitors such as vemurafenib and dabrafenib. Although it has been demonstrated that these targeted drugs significantly improve PFS and OS in comparison with chemotherapy, the patients receiving this treatment rapidly develop resistance [2]. On the other hand, the FDA-approved checkpoint inhibitors against cytotoxic T lymphocyte antigen 4 (CTLA-4) and programmed death 1 (PD-1) enhance the natural antitumor immune response of the patients and also lead to improved survival [18]. However, only a subset of patients respond to immune checkpoint inhibitors and resistance mechanisms can also arise among this group of responders. In this context, the identification of predictive biomarkers and/or baseline covariates able

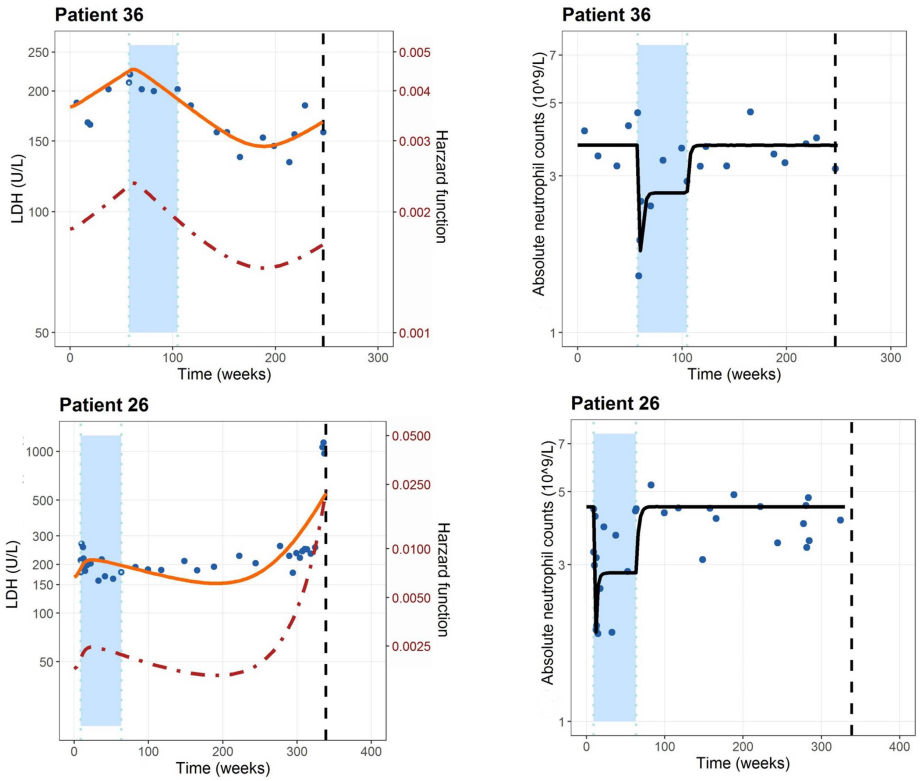


Fig. 4.6 Individual predicted LDH and ANC profiles (solid lines) by the selected models and the time course of the hazard rate (dashed line) differentiating by an individual who is alive at the last follow-up (patient 36) and an individual who died (patient 26). Solid points represents the observation values of the patients.

Table 4.4 Median value (5–95th) of different outcomes obtained after 1000 simulation of different dosing schedules.

	1-year	2-year	5-year	10-year
Standard dosing regimen*: induction: 20 MU/m^2 5 days a week + maintenance: 10 MU/m^2 3 days a week				
LDHrel	-0.127 (-0.29-0.095)	-0.242 (-0.52-.019)	-0.062 (-0.61-1)	0.138 (-0.6-1.79)
Overall Survival	91.5 (89.7-93.2)	83.6 (81.3-85.9)	58.5 (55.5-61.6)	29.5 (26.8-32.4)
Absolut neutrophil count ($10^9/L$)	2.14 (1.04-4.07)	3.21 (1.94-5.36)	3.17 (1.93-5.66)	3.92 (2.37-6.15)
Regimen 2*: induction: 40 MU/m^2 5 days a week + maintenance: 10 MU/m^2 3 days a week				
LDHrel	-0.133 (-0.3-0.088)	-0.247 (-0.53-.018)	-0.07 (-0.62-0.99)	0.129 (-0.619-1.74)
Overall Survival	91.6 (89.9-93.3)	83.7 (81.4-86)	58.8 (55.8-61.9)	29.7 (27-32.7)
Absolut neutrophil count ($10^9/L$)	2.14 (1.03-4.07)	3.21 (1.94-5.36)	3.17 (1.93-5.66)	3.92 (2.36-6.14)
Regimen 3*: induction: 20 MU/m^2 5 days a week + maintenance: 20 MU/m^2 3 days a week				
LDHrel	-0.132 (-0.3-0.09)	-0.266 (-0.53-.016)	-0.1 (-0.66-0.94)	0.134 (-0.66-1.62)
Overall Survival	91.5 (89.8-93.2)	83.6 (81.3-86)	59.8 (56.8-62.9)	31.5 (28.7-34.5)
Absolut neutrophil count ($10^9/L$)	1.4 (0.4-3.27)	3.21 (1.93-5.36)	3.15 (1.93-5.66)	3.87 (2.3-6.1)

*The treatment started 6 weeks after diagnosis for all the regimens simulated. The treatment was administered to an individual with a body surface area of 1.8 m^2. LDHrel is the relative change of LDH from the baseline value.

to select patients most likely to benefit from these therapies could be crucial.

In our case, none of the studied baseline covariates (Breslow thickness, tumor extension, presence of ulceration...) influenced survival. Unexpectedly, this agrees with the result of other statistical analysis made in advanced melanoma patient data where the univariate analysis of the gender, age, Breslow thickness, BRAF mutation status and location of primary tumor resulted in no significant association with OS [11, 16]. Still, we find difficult to conclude that these covariates do not influence the survival of melanoma patients because we think that part of the results obtained were influenced by the small number of patients in the dataset and the missing information regarding the covariates of these subjects. Therefore, a better univariate and multivariate baseline covariate analysis is encouraged if more informative datasets are available in the future. Other limitations to highlight regarding this work were that no PK and tumor progression measurements were available for the development of the model. That is why a K-PD approach was used to link the dosing records with a drug effect in an unobserved variable that simulates the disease progression of the patients. This tumor progression was in turn linked to the time course of LDH, MIA and S100B serum concentration dynamics that were produced by a first-order rate constant and cleared at a first-order elimination rate in healthy subjects. However, the major obstacle to develop this treatment-biomarker-survival-toxicity framework to monitor clinical response in melanoma was the moderate efficacy of IFN-$\alpha 2b$ therapy. Although the 1684 ECOG trial probed a significant improvement on the PFS and OS of melanoma patients, subsequent trials showed limited efficacy of this treatment as monotherapy, particularly on the OS of the individuals [12]. Table 4 shows how doubling the dose from the induction or the maintenance phase of the treatment influenced the LDH values, OS and ANC of 1000 simulated individuals for a 1, 2, 5 and 10 year period. The values summarized in this table showed that doubling the dose of the induction or maintenance phases doesn't have much repercussion in OS due to the low drug effects, but altering the maintenance phase could provoke a lower neutropenia grade. Even so, the development of new treatments opens the opportunity to reanalyze the utility of these tumor markers as prognostic factors (in fact, LDH has been reported as a clinically significant factor associated with OS under targeted and immune therapies [11, 16]) and to follow-up patients during therapy [19] re-using parts of the model built in this work. More importantly, the modeling effort developed here offers an attractive methodology to evaluate not only new treatment alternatives in drug development but also existing ones in the clinic, in order to evaluate safety and efficacy of the therapy, identify predictive factors and biomarkers and finally, perform dosing optimization in order to improve the clinical outcome of the patients.

References

[1] Akaike, H. (1987). Factor analysis and AIC. In *Selected Papers of Hirotugu Akaike*, pages 371–386. Springer.

[2] Azijli, K., Stelloo, E., Peters, G. J., and VAN DEN Eertwegh, A. J. M. (2014). New developments in the treatment of metastatic melanoma: immune checkpoint inhibitors and targeted therapies. *Anticancer Res.*, 34(4):1493–1505.

[3] Bergstrand, M., Hooker, A. C., Wallin, J. E., and Karlsson, M. O. (2011). Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J.*, 13(2):143–151.

[4] Buil-Bruna, N., López-Picazo, J.-M., Moreno-Jiménez, M., Martín-Algarra, S., Ribba, B., and Trocóniz, I. F. (2014). A population pharmacodynamic model for lactate dehydrogenase and neuron specific enolase to predict tumor progression in small cell lung cancer patients. *AAPS J.*, 16(3):609–619.

[5] Desmée, S., Mentré, F., Veyrat-Follet, C., and Guedj, J. (2015). Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: A comparison by simulation of two-stage and joint approaches. *AAPS J.*, 17(3):691–699.

[6] Desmée, S., Mentré, F., Veyrat-Follet, C., Sébastien, B., and Guedj, J. (2017). Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear PSA kinetics and survival in prostate cancer patients. *Biometrics*, 73(1):305–312.

[7] Friberg, L. E., Henningsson, A., Maas, H., Nguyen, L., and Karlsson, M. O. (2002). Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *J. Clin. Oncol.*, 20(24):4713–4721.

[8] Garbe, C., Peris, K., Hauschild, A., Saiag, P., Middleton, M., Bastholt, L., Grob, J.-J., Malvehy, J., Newton-Bishop, J., Stratigos, A. J., Pehamberger, H., Eggermont, A. M., European Dermatology Forum (EDF), European Association of Dermato-Oncology (EADO), and European Organisation for Research and Treatment of Cancer (EORTC) (2016). Diagnosis and treatment of melanoma. european consensus-based interdisciplinary guideline - update 2016. *Eur. J. Cancer*, 63:201–217.

[9] Gray, M. R., Martin del Campo, S., Zhang, X., Zhang, H., Souza, F. F., Carson, 3rd, W. E., and Smith, A. D. (2014). Metastatic melanoma: lactate dehydrogenase levels and CT imaging findings of tumor devascularization allow accurate prediction of survival in patients treated with bevacizumab. *Radiology*, 270(2):425–434.

[10] Jacqmin, P., Snoeck, E., van Schaick, E. A., Gieschke, R., Pillai, P., Steimer, J.-L., and Girard, P. (2007). Modelling response time profiles in the absence of drug concentrations: Definition and performance evaluation of the K–PD model. *J. Pharmacokinet. Pharmacodyn.*, 34(1):57–85.

[11] Kelderman, S., Heemskerk, B., van Tinteren, H., van den Brom, R. R. H., Hospers, G. A. P., van den Eertwegh, A. J. M., Kapiteijn, E. W., de Groot, J. W. B., Soetekouw, P., Jansen, R. L., Fiets, E., Furness, A. J. S., Renn, A., Krzystanek, M., Szallasi, Z., Lorigan, P., Gore, M. E., Schumacher, T. N. M., Haanen, J. B. A. G., Larkin, J. M. G., and Blank, C. U. (2014). Lactate dehydrogenase as a selection criterion for ipilimumab treatment in metastatic melanoma. *Cancer Immunol. Immunother.*, 63(5):449–458.

[12] Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M., and Blum, R. H. (2000). High- and low-dose interferon alfa-2b in high-risk melanoma: first analysis of intergroup trial E1690/S9111/C9190. *J. Clin. Oncol.*, 18(12):2444–2458.

[13] Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C., and Blum, R. H. (1996). Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the eastern cooperative oncology group trial EST 1684. *J. Clin. Oncol.*, 14(1):7–17.

[14] Lindbom, L., Pihlgren, P., and Jonsson, N. (2005). PsN-Toolkit—A collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput. Methods Programs Biomed.*, 79(3):241–257.

[15] Lindbom, L., Ribbing, J., and Jonsson, E. N. (2004). Perl-speaks-NONMEM (PsN)–a perl module for NONMEM related programming. *Comput. Methods Programs Biomed.*, 75(2):85–94.

[16] Long, G. V., Grob, J.-J., Nathan, P., Ribas, A., Robert, C., Schadendorf, D., Lane, S. R., Mak, C., Legenne, P., Flaherty, K. T., and Davies, M. A. (2016). Factors predictive of response, disease progression, and overall survival after dabrafenib and trametinib combination treatment: a pooled analysis of individual patient data from randomised trials. *Lancet Oncol.*, 17(12):1743–1754.

[17] Palmer, S. R., Erickson, L. A., Ichetovkin, I., Knauer, D. J., and Markovic, S. N. (2011). Circulating serologic and molecular biomarkers in malignant melanoma. *Mayo Clin. Proc.*, 86(10):981–990.

[18] Rausch, M. P. and Hastings, K. T. (2018). Immune checkpoint inhibitors in the treatment of melanoma: From basic science to clinical application. In Ward, W. H. and Farma, J. M., editors, *Cutaneous Melanoma: Etiology and Therapy*. Codon Publications, Brisbane (AU).

[19] Sanmamed, M. F., Fernández-Landázuri, S., Rodríguez, C., Lozano, M. D., Echeveste, J. I., Pérez Gracia, J. L., Alegre, E., Carranza, O., Zubiri, L., Martín-Algarra, S., and González, A. (2014). Relevance of MIA and S100 serum tumor markers to monitor BRAF inhibitor therapy in metastatic melanoma patients. *Clin. Chim. Acta*, 429:168–174.

[20] Tang, M., Zhao, R., van de Velde, H., Tross, J. G., Mitsiades, C., Viselli, S., Neuwirth, R., Esseltine, D.-L., Anderson, K., Ghobrial, I. M., San Miguel, J. F., Richardson, P. G., Tomasson,

M. H., and Michor, F. (2016). Myeloma cell dynamics in response to treatment supports a model of hierarchical differentiation and clonal evolution. *Clin. Cancer Res.*, 22(16):4206–4214.

[21] Zhang, L., Beal, S. L., and Sheiner, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data i: best-case performance. *J. Pharmacokinet. Pharmacodyn.*, 30(6):387–404.

Supplementary Figures

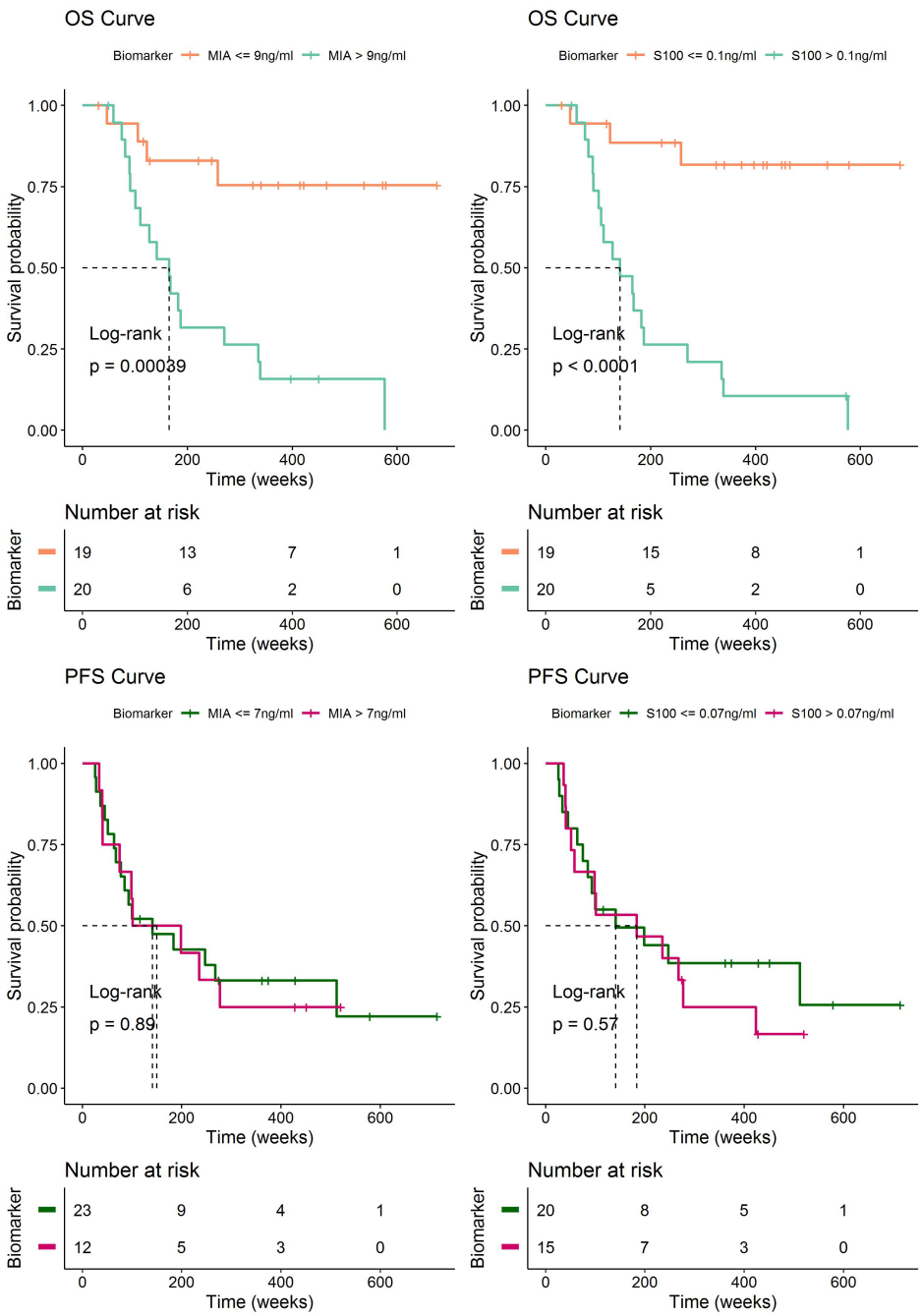


Fig. S4.1 Evaluation of the overall survival (OS) and progression-free survival (PFS) of the patients with high and low MIA and S100B concentrations at the end of the study.

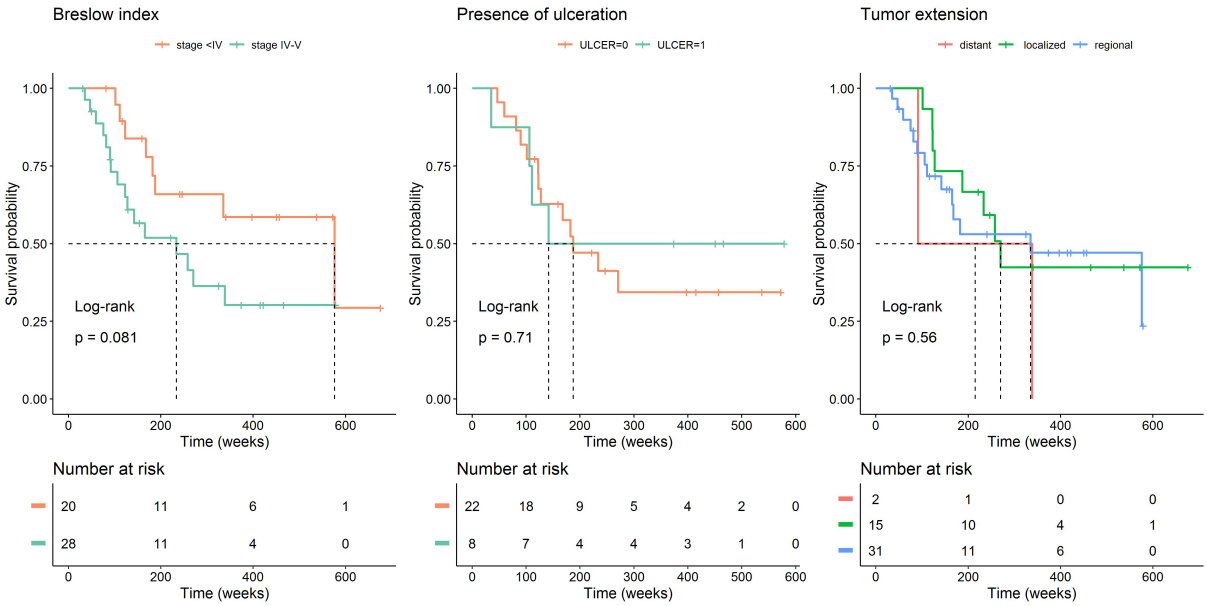


Fig. S4.2 Kaplan-Meier curves of the OS stratified by different baseline covariates.

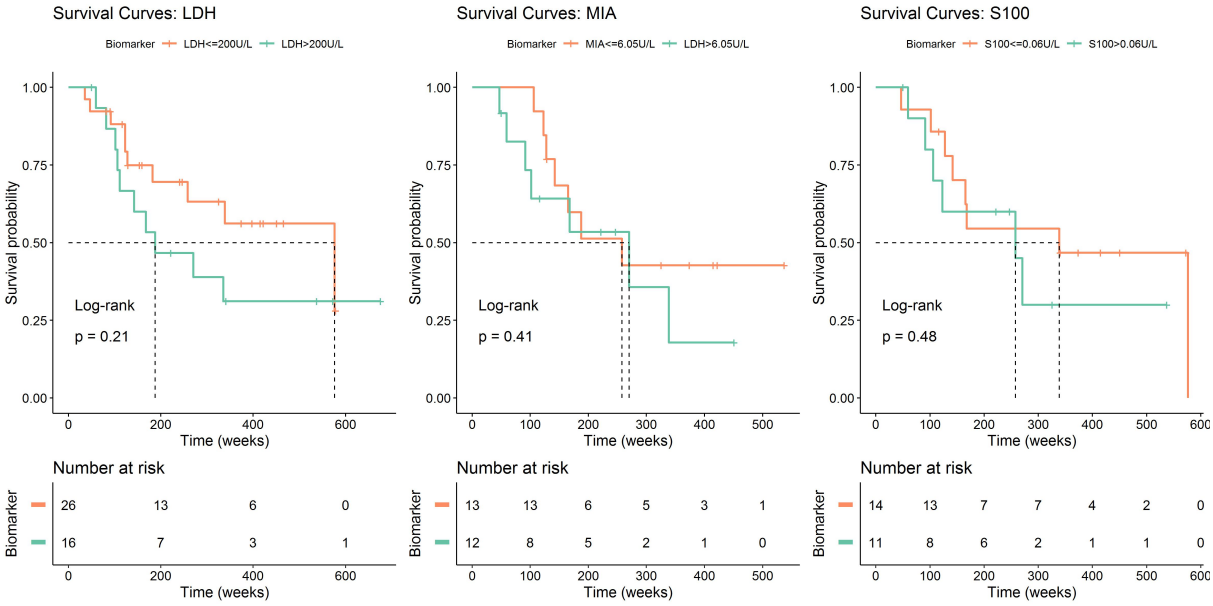


Fig. S4.3 Kaplan-Meier curves of the OS stratified by baseline biomarker values (before treatment initiation).

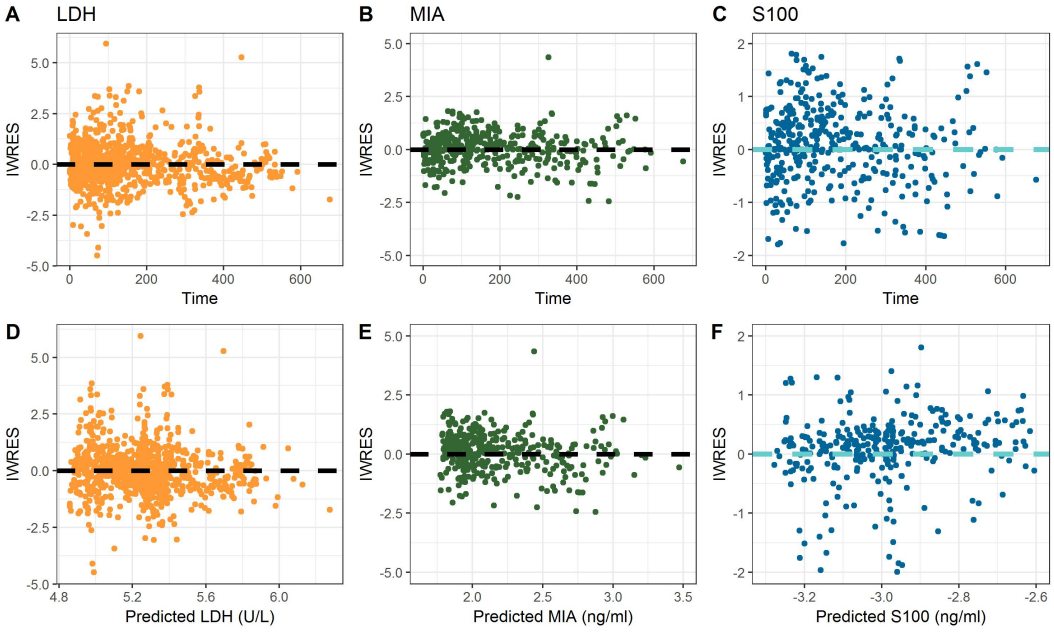


Fig. S4.4 Individual Weigthed RESiduals (IWRES) plot.

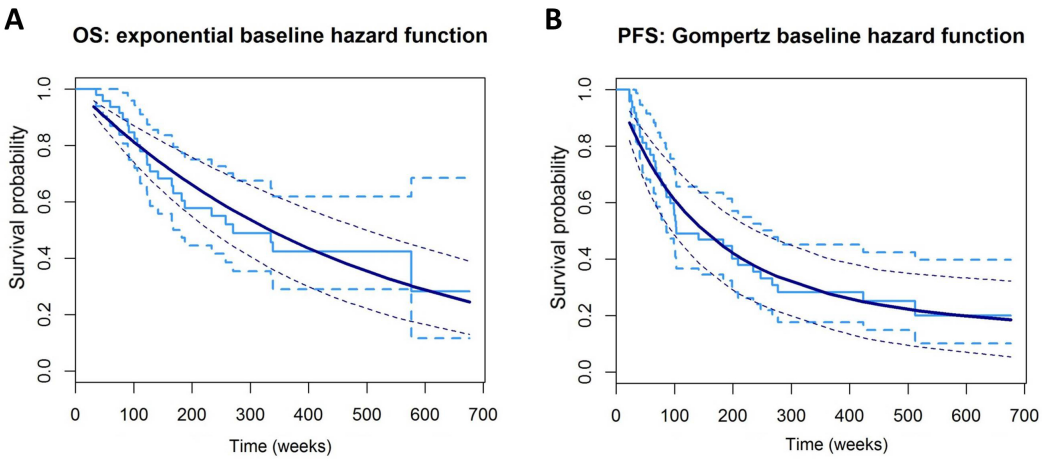


Fig. S4.5 Best fitted models for the baseline hazard function of the overall survival (OS) and progression-free survival (PFS) of the patients

General discussion

The discipline of Pharmacometrics and Systems Pharmacology (PSP) is focused on the development of pharmacokinetic (PK), pharmacodynamics (PD) and disease progression models to interpret, describe and predict drug effects in the target patient populations supporting key decisions during all phases of drug discovery and development, and patient management through personalize medicine including dose individualization. As it has become clear through the different chapters of this thesis, PSP integrates principles from the field of biology, pharmacology, biostatistics, mathematics and computer science for proper use of available information in order to get a better understanding of the in vivo drug effects and the variability associated to treatment. These mathematical models provide deeper insights related to emergent pharmacological properties that could remain hidden in the raw data following empirical and standard statistical approaches, thus supporting hypothesis generation and challenging model predictions against experimental data. It is therefore not surprising that pharma industry, regulatory agencies as well as academia and clinics consider PSP as a relevant key player in their corresponding arenas.

In general, modeling approaches can be classified in different groups depending on the mathematical formalisms they are based on. Models can be quantitative when mathematical equations are used to describe the relationship between the components of the systems, or qualitative when the relationship between the model components is based on signed and directed causal regulatory graphs. Another possible classification discussed in this thesis distinguishes between deterministic and stochastic models. In deterministic modeling, stochasticity within the system is neglected and the output of the model is fully determined by the parameter values and the initial conditions. Stochastic models on the other hand, have some inherent randomness, that is, the same set of parameter values and initial conditions can lead to different results. The two classifications above, namely qualitative/quantitative and deterministic/stochastic, are complementary, that is, a quantitative or qualitative model can be either deterministic or stochastic.

A key factor driving the choice between qualitative and quantitative representations is the type of knowledge and information available related to the system under study. When there is no longitudinal experimental data available but connections between the components of a system are known at least from a qualitative perspective, a network-based approach might be a suitable option [10].
The next question to answer then is how to build this type of networks. Depending on the system being modelled, a bottom-up or knowledge-based approach can be adopted, where information about the components to include and their relationships is gathered from scientific literature or public repositories that contain previously generated models. In fact, both the Boolean network examples annexed in this thesis were constructed using this bottom-up approach. By contrast, it is sometimes possible to infer the topology of the networks directly from experimental datasets (e.g. gene expression data [3]), which is known as a top-down or data-based approach. Data-based methods allow the construction of directed graphs whose edges represent positive or negative effects of the components under study.

Those graphs can then be transformed to logic models using specific software like *CellNetOptimizer* [7]. Although choosing this procedure may seem less painful for the researcher a priori, the networks obtained with the top-down approach also need an extensive manual curation.

As an example of qualitative strategy, in *Chapter 1* we proposed the use of Boolean models (i.e. discrete two-state logical models in which each node in a network is represented as a simple on/off switch) to analyze the results of SP networks. Here we showed that the updating method of the nodes in the network (synchronous vs asynchronous) might have a big influence in the outcome of the model. In this sense, in the synchronous updating method the output of the network will always be the same when starting from an initial condition whereas with the asynchronous approach different outputs could be obtained. These characteristics correspond to the definition of deterministic and stochastic modeling strategies previously mentioned and therefore here we demonstrated how a qualitative framework can also have a deterministic or stochastic behavior depending on the chosen algorithm.

Currently the application of Boolean analysis to PSP is still very limited, contrary to the case of applying deterministic models to continuous data, where there is a battery of tools to help the scientist for model implementation, fitting and evaluation. We think that with the development of SPIDDOR we provide and easy-to-use tool to perform simulations and analyze the results of previously constructed Boolean networks (using the bottom-up or the top-down approach).

Once a Boolean model is built, one can produce trajectories and study the possible attractors of the system (i.e. steady-states). Especially when the networks are highly interconnected, analyzing the state-transition diagrams of the attractors is very difficult because they generally consist on complex cycles. This analysis is further complicated when node perturbations are introduced on the network (knockouts, overexpressions...) and the differences obtained between the perturbed and the normal state of the network want to be determined. In SPIDDOR, we simplified this analysis by representing the activation probability of the nodes inside the attractors. This and other advantages of the framework have been already discussed in *Chapter 1*.

In addition, in the Appendix section we also showed two applications where SPIDDOR has been used to simulate Boolean networks related to autoimmune diseases. In [9] the potential of a Boolean network model to support drug development in systemic lupus erythematosus was explored, focusing the analysis on the antigen presentation to naive T cells. In [1] on the other hand, a logic model for inflammatory bowel disease is proposed to describe the pathogenic mechanisms of the disorder and qualitatively describe the characteristic chronic inflammation.

In these works we showed that logic models are a useful tool to make a general picture of the systems under study and test potential therapeutic interventions. It must be addressed however, that in this type of networks there is no a clear notion of continuous time and therefore, as longitudinal data becomes

available, the transformation of these networks to a more quantitative framework is essential.

In this sense, non-linear mixed-effects (NLME) modelling has proven to be one of the most successful data-driven strategy to characterize longitudinal data coming from different individuals. NLME approach is widely used in PSP in order to quantitatively characterize the time course of disease progression and establish the link to the PKPD properties of the drugs. As such, PKPD modeling constitutes the scientific basis for dose and delivery optimization of new and existing drugs.

It is our perception however that these capabilities of the models are far from being fully exploited and the question of "how and to which degree the system needs to be perturbed (optimized) to achieve certain physiological or therapeutic conditions" is in general not addressed. This is mainly due to the fact that (i) under complex dynamic systems and multiple therapeutic objectives to achieve, optimization cannot be efficiently addressed through trial-error simulation exercises, and (ii) there is a lack of standard methodology to implement formal optimization approaches in computational biology.

Indeed, in *Chapter 2*, we investigated how Optimal Control (OC) principles could be applied to derive optimal drug delivery profiles of the drug triptorelin, a gonadotropin-releasing hormone analog, using a PKPD model for its testosterone (TST) effects in prostate cancer patients [8]. The aim of this work was to derive the optimal drug release profiles to achieve the following multiple therapeutic goals: minimize both the initial flare up of TST levels and the time to reach TST values below castration limit (t_{cast}), while maximizing the castration period of the patients (for more than 9 months). As the value of t_{cast} was not known in advance, the problem was divided into two phases each represented by a different cost function and constraints and defined between: (i) [0, t_{cast}] and (ii) [t_{cast}, 280+t_{cast}] days, respectively. Once the optimal TST profiles were obtained, we directly approximate them using the zero-order release compartment and two first-order release compartments from the original PKPD model and estimate the most adequate absorption parameters using NLME methods.

This work presented a good example of complex dynamic system with multiple therapeutic objectives to achieve where a more advanced optimization technique coming from engineering has shown highly favorable results. In the *Introduction* of this thesis and in the discussion from *Chapter 2* we also indicated that the proposed approach is not circumscribed to just this particular problem, but also to many other problems related to drug exposure optimization. Thus optimal control is a promising multidisciplinary modeling technique that not only could help us to improve patient coverage but also to increase drug development successes.

On the other hand, in *Chapter 3* the disciplines of NLME and stochastic models have been combined in order to create a more comprehensive and practical computational framework called ACESO (A

Cancer Evolution Simulation Optimizer) aimed to explore the dynamic evolution of heterogeneous tumor cell populations while taking pharmacokinetics and drug interaction effects into account. This approach is made up of a cell-level description of the changes in drug sensitive and resistant cell populations over time and in response to treatment in the form of a stochastic model known as multitype branching process. In this model, sensitive cells accumulate mutations at a given rate per cell division, generating new clones harboring specific resistance mechanisms to the drugs. The birth and death rate of each cell type are influenced by drug concentrations, and hence, accounting for population pharmacokinetic models is essential to form a multiscale description of drug pharmacokinetics/pharmacodynamics and cancer evolution. The ultimate goal of this tool is to search through different possible drug administration strategies to identify the one that is predicted to be best, for instance, because it minimizes the risk of resistance or the expected number of cancer cells over time.

We must note however that, although the framework of this work is based on a stochastic model of clonal evolution, the equations implemented in ACESO are deterministic, as only the mathematical approximations for the expected values of the total number of cells and the probability of developing resistance have been incorporated in the package in order to accelerate the process of searching for the optimal dosing schedule. Even if the user is able to add variability to the PK parameters and test different initial conditions and birth, death and mutation rate constants, the possibility of using a stochastic simulation algorithm (SSA) to analyze the evolution of the cells is not implemented.

Still, the utility of this tool has been demonstrated through different case studies mainly focused on the analysis of the evolution of different populations of tumor cells under targeted therapies. Nonetheless, the tool is very versatile and could also be used to study the effects of other type of anti-cancer treatments or other biological problems like the emergence of antibiotic resistance in bacteria.

We think that the activity of combining models could become a major part of PSP in near future. However, an issue regarding the combination of different methodologies is that the different frameworks may require very different numerical approaches towards equations integration. These problems could arise when combining ordinary differential equations with stochastic kinetics (represented as master equations or stochastic differential equations) or qualitative (e.g. Boolean networks) and quantitative models (e.g. ODEs). In the former, some examples already exist where the combination of SDE and ODE has been implemented in NONMEM [12]. However, for the case of stochastic models based on master equations, mathematical approximations of the processes being modeled is encouraged in order to easily integrate them in ODE-based systems. Although it is possible to represent part of a model with differential equations and another part with a stochastic simulation algorithm once a parametrization of the model has been done, the integrated model will be much more computationally intensive. In addition, the task of coding a SSA will be quite complex for researchers without a computer science background. As better techniques become available that could

facilitate the parametrization and simulation of stochastic models, frameworks like ACESO make the integration of models coming from different disciplines easier.

Regarding the integration of qualitative and quantitative modeling approaches, methods exist for the conversion of Boolean systems into systems of ordinary differential equations by multivariate polynomial interpolation or the application of sigmoidal Hill functions [6]. In contrast to these deterministic logic-based ODEs, *MaBoSS* software simulates continuous time Markov processes on Boolean networks to handle the asynchronous updates of the states of the nodes in a stochastic way and generate a population of trajectories as sequences of Boolean states [11]. Transition rates can be associated with each node, and probabilities of network states can be estimated given a set of initial probabilities.

However, when the reason of initially choosing a logic-based approach is the insufficient longitudinal data available to characterize the biological process of interest, the quantitative version of the Boolean model will be difficult to calibrate. Another concern is with regard to the size and complexity of those models. ODE based systems with a large number of components have a vast number of parameters, which results in a large uncertainty in parameter estimates even if some of them are fixed from values obtained from literature. Parameter estimation in stochastic models is even more difficult and is not yet adequately addressed in the literature.

While there are still many aspects that need to be improved, we strongly believe that the next generation of models will be more modular, in which different processes could be modelled independently and then integrated in the simulation process through variable transformation and synchronization. Certainly, this integration requires an additional effort from scientists and engineers who have to become familiar with these approaches and simplify the work for the next generations.

Lastly, in the chapters corresponding to this dissertation, not only different type of methodologies have been explored but also the combination of models of the same nature that together are able to characterize and integrate different type of data. As an example, the framework presented in *Chapter 4* describes the relationship between the time-course of three biomarkers (longitudinal data) and the overall survival of melanoma patients (time-to-event data) while taking the side effects of the treatment into account. The use of this type of models is not innovative in PSP; recent examples can be found in [2, 5, 4]. However, due to the increasing interest in the use of biomarkers in drug development and clinical practice in order to recruit patients with the most favorable benefit/risk ratio, we also found interesting to include this analysis as the last chapter of the dissertation.

We conclude that the different modeling strategies discussed in this thesis may help in the development of more mechanistic pharmacological modelling approaches that could reduce the late-stage attrition in the drug development process, but also in the transition from drug development to patient care in order to drive optimal intervention strategies resulting from the combination of individual physiological information and computational methodologies.

References

[1] Balbas-Martinez, V., Ruiz-Cerdá, L., Irurzun-Arana, I., González-García, I., Vermeulen, A., Gómez-Mantilla, J. D., and Trocóniz, I. F. (2018). A systems pharmacology model for inflammatory bowel disease. *PLoS One*, 13(3):e0192949.

[2] Buil-Bruna, N., López-Picazo, J.-M., Moreno-Jiménez, M., Martín-Algarra, S., Ribba, B., and Trocóniz, I. F. (2014). A population pharmacodynamic model for lactate dehydrogenase and neuron specific enolase to predict tumor progression in small cell lung cancer patients. *AAPS J.*, 16(3):609–619.

[3] Clarke, D. J. B., Kuleshov, M. V., Schilder, B. M., Torre, D., Duffy, M. E., Keenan, A. B., Lachmann, A., Feldmann, A. S., Gundersen, G. W., Silverstein, M. C., Wang, Z., and Ma'ayan, A. (2018). eXpression2Kinases (X2K) web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res.*, 46(W1):W171–W179.

[4] Desmée, S., Mentré, F., Veyrat-Follet, C., Sébastien, B., and Guedj, J. (2017). Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear PSA kinetics and survival in prostate cancer patients. *Biometrics*, 73(1):305–312.

[5] Garcia-Cremades, M., Pitou, C., Iversen, P. W., and Troconiz, I. F. (2018). Predicting tumour growth and its impact on survival in gemcitabine-treated patients with advanced pancreatic cancer. *Eur. J. Pharm. Sci.*, 115:296–303.

[6] Krumsiek, J., Pölsterl, S., Wittmann, D. M., and Theis, F. J. (2010). Odefy–from discrete to continuous models. *BMC Bioinformatics*, 11:233.

[7] Morris, M. K., Melas, I., and Saez-Rodriguez, J. (2012). Construction of cell type-specific logic models of signalling networks using CellNetOptimizer. *Methods in Molecular Biology: Computational Toxicology, Ed. B. Reisfeld and A. Mayeno, Humana Press*.

[8] Romero, E., Vélez de Mendizabal, N., Cendrós, J.-M., Peraire, C., Bascompta, E., Obach, R., and Trocóniz, I. F. (2012). Pharmacokinetic/pharmacodynamic model of the testosterone effects of triptorelin administered in sustained release formulations in patients with prostate cancer. *J. Pharmacol. Exp. Ther.*, 342(3):788–798.

[9] Ruiz-Cerdá, M. L., Irurzun-Arana, I., González-Garcia, I., Hu, C., Zhou, H., Vermeulen, A., Trocóniz, I. F., and Gómez-Mantilla, J. D. (2016). Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. *Eur. J. Pharm. Sci.*

[10] Sorger, P. K., Allerheiligen, S. R. B., Abernethy, D. R., Altman, R. B., Brouwer, K. L. R., Califano, A., D'Argenio, D. Z., Iyengar, R., Jusko, W. J., Lalonde, R., and Others (2011). Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs

and understanding therapeutic mechanisms. In *An NIH white paper by the QSP workshop group*, volume 48.

[11] Stoll, G., Caron, B., Viara, E., Dugourd, A., Zinovyev, A., Naldi, A., Kroemer, G., Barillot, E., and Calzone, L. (2017). MaBoSS 2.0: an environment for stochastic boolean modeling. *Bioinformatics*, 33(14):2226–2228.

[12] Tornøe, C. W., Overgaard, R. V., Agersø, H., Nielsen, H. A., Madsen, H., and Jonsson, E. N. (2005). Stochastic differential equations in NONMEM: implementation, application, and comparison with ordinary differential equations. *Pharm. Res.*, 22(8):1247–1258.

Conclusions/Conclusiones

The primary goal of the present project was to explore the use of different methodologies and tools that could benefit the field of Pharmacometrics and Systems Pharmacology modelling.

Several achievements have been already summarized in the different chapters of this dissertation. These contributions include:

1. A versatile tool called SPIDDOR, which is specifically tailored to the design and analysis of Boolean network models in the area of Systems Pharmacology. This tool proposes several novelties, including:

    - new types of regulatory interactions, the positive and negative modulations, which lead to richer dynamics between the nodes.
    - Improvement of the perturbation paradigm emulating a "polymorphism" of a node.
    - Improvement of the exploratory analysis of the output of the simulations computed on these models by incorporating i) new visualization techniques to evaluate the attractors (long-time behavior) of the system and the effects of perturbations and ii) a clustering method to group the nodes that lead to similar alterations within the network.

2. Two boolean network models describing pathogenic mechanisms in the autoimmune diseases systemic lupus erythematosus and inflammatory bowel disease built and analyzed with SPIDDOR (*Appendix* section).

3. The application of optimal control theory to a population pharmacokinetic/pharmacodynamic model for the testosterone effects of triptorelin in prostate cancer patients in order to derive the optimal drug release profile to achieve multiple therapeutic goals. Apart from highlighting the advantages of using this optimization technique to define a reverse engineering problem, the information summarized in this work could be very useful for the development of new formulations, since it provides insight into the desired absorption characteristics that could produce a broad benefit for future patients.

4. A multi-scale computational framework in R called ACESO (A Cancer Evolution Simulation Optimizer) which incorporates a multitype branching process model where the birth, death and mutation rates of the heterogeneous tumor cell populations can be influenced by the drug concentrations resulting from the simulation of pharmacokinetic models. This approach is an important step towards validating evolutionary mathematical models in the clinic as it investigates the evolution of resistant clones to anti-cancer therapies while taking the pharmacokinetic

of the drugs and drug interaction effects into account.

5. A semi-mechanistic model describing the time course of several circulating biomarkers in advanced melanoma patients treated with adjuvant high-dose interferon-alpha2b to evaluate the dynamics of the tumor markers as prognostic factors of the overall survival and progression-free survival of the patients. A semi-mechanistic myelosuppression model to evaluate the adverse effects of the therapy has been also developed in order to convert the individual biomarker levels into personalized predictions of survival while taking toxicity into account.

6. Apart from the underlying methodology exposed in each chapter and the corresponding publications, this thesis also involved the development of various tools, two R packages especifically, that allows for efficient use of some of these methods, demonstrating their usefulness through the different case studies presented in each chapter and appendix section.

El objetivo principal del presente proyecto ha sido el de explorar el uso de diferentes metodologías y herramientas que puedan beneficiar el campo de la farmacometría y la farmacología de sistemas.

Los distintos logros han sido expuestos en los diferentes capítulos de esta disertación. Estas contribuciones incluyen:

1. Una versátil herramienta denominada SPIDDOR, específicamente diseñada para el desarrollo y análisis de modelos de redes booleanas en el área de farmacología de sistemas. Esta herramienta propone varias novedades, entre ellas:
    - Nuevos tipos de interacciones reguladoras, las modulaciones positivas y negativas, que conducen a dinámias más diversas entre los nodos del sistema.
    - Una mejora en el paradigma de las perturbaciones de estas redes emulando la acción de "polimorfismos" en los nodos.
    - Mejora del análisis exploratorio de las simulaciones obtenidas a partir de estos modelos mediante la incorporación de i) nuevas técnicas de visualización de los atractores (comportamiento a largo plazo) del sistema y los efectos de las perturbaciones y ii) métodos de agrupamiento ("*clustering*") para asociar los nodos que producen alteraciones similares dentro de la red.

2. Dos modelos de redes booleanas analizados utilizando SPIDDOR que describen mecanismos patogénicos en las enfermedades autoinmunes lupus eritematoso sistémico y la enfermedad inflamatoria intestinal (ver Apéndice).

3. La aplicación de técnicas de control óptimo ("*Optimal Control*") a un modelo farmacocinético/farmacodinámico poblacional que describe los efectos de la triptorelina en la testosterona de los pacientes con cáncer de próstata con el fin de obtener los perfiles de liberación del fármaco óptimos que logren múltiples objetivos terapéuticos. Además de resaltar las ventajas de usar esta técnica de optimización para definir un problema de ingeniería inversa, la información resumida en este trabajo podría ser muy útil para el desarrollo de nuevas formulaciones, ya que brinda información sobre las características de absorción deseadas que podrían generar un amplio beneficio para futuros pacientes.

4. Una herramienta computacional basada en el entorno R llamada ACESO (A Cancer Evolution Simulation Optimizer) que incorpora un modelo de proceso de ramificación de múltiples tipos ("*multitype branching process*") en el que las tasas de división, muerte y mutación de las poblaciones de células tumorales heterogéneas pueden verse influenciadas por las concentraciones

de fármaco resultantes de la simulación de modelos farmacocinéticos. Este enfoque es un paso importante hacia la validación de modelos matemáticos evolutivos en la clínica, ya que investiga la evolución de los clones resistentes a las terapias contra el cáncer a la vez que toma en cuenta la farmacocinética de los medicamentos y los efectos de interacción de los mismos.

5. Un modelo semi-mecanístico que describe el curso temporal de varios biomarcadores circulantes en pacientes con melanoma avanzado tratados con dosis altas de interferón-alfa2b adyuvante para evaluar la dinámica de estos marcadores tumorales como posibles factores pronóstico de la supervivencia general y la supervivencia libre de progresión de los pacientes. Además, se ha desarrollado un modelo de mielosupresión para evaluar los efectos adversos de la terapia a fin de convertir los niveles de biomarcador individuales en predicciones personalizadas de supervivencia mientras se tiene en cuenta la toxicidad de la terapia.

6. Además de la metodología expuesta en cada capítulo y las publicaciones correspondientes, este trabajo ha involucrado el desarrollo de varias herramientas, dos paquetes de R específicamente, que permiten el uso eficiente de algunos de los métodos expuestos, demostrando su utilidad a través de los diferentes ejemplos presentados en cada capítulo y el apéndice de esta tesis.

Appendix

A

Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach

Ruiz-Cerdá ML, Irurzun-Arana I, González-Garcia I, Hu C, Zhou H, Vermeulen A, Trocóniz IF, Gómez-Mantilla JD. Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. Eur J Pharm Sci. 2016, 94:46-58. http://doi.org/10.1016/j.ejps.2016.04.010

B

A Systems Pharmacology Model for Inflammatory Bowel Disease

PLOS ONE

RESEARCH ARTICLE

A systems pharmacology model for inflammatory bowel disease

Violeta Balbas-Martinez[1,2], Leire Ruiz-Cerdá[1,2], Itziar Irurzun-Arana[1,2], Ignacio González-García[1¤a], An Vermeulen[3,4], José David Gómez-Mantilla[1¤b], Iñaki F. Trocóniz[1,2]*

1 Pharmacometrics & Systems Pharmacology, Department of Pharmacy and Pharmaceutical Technology, School of Pharmacy and Nutrition, University of Navarra, Pamplona, Spain, **2** IdiSNA, Navarra Institute for Health Research, Pamplona, Spain, **3** Janssen Research and Development, a division of Janssen Pharmaceutical NV, Beerse, Belgium, **4** Laboratory of Medical Biochemistry and Clinical Analysis, Faculty of Pharmaceutical Sciences, Ghent, Belgium

¤a Current address: PharmaMar, Colmenar Viejo, Madrid, Spain.
¤b Current address: Boehringer Ingelheim, Ingelheim am Rhein, Germany.
* itroconiz@unav.es

OPEN ACCESS

Citation: Balbas-Martinez V, Ruiz-Cerdá L, Irurzun-Arana I, González-García I, Vermeulen A, Gómez-Mantilla JD, et al. (2018) A systems pharmacology model for inflammatory bowel disease. PLoS ONE 13(3): e0192949. https://doi.org/10.1371/journal.pone.0192949

Editor: Shree Ram Singh, National Cancer Institute, UNITED STATES

Received: October 17, 2017

Accepted: February 1, 2018

Published: March 7, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Development of the computational model was supported by a fellowship grant from the Navarra Government to Violeta Balbás-Martínez of 61.965 Euros (http://www.navarra.es/home_es/Actualidad/BON/Boletines/2017/18/Anuncio-5/) and Janssen Research and Development. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of

Abstract

Motivation

The literature on complex diseases is abundant but not always quantitative. This is particularly so for Inflammatory Bowel Disease (IBD), where many molecular pathways are qualitatively well described but this information cannot be used in traditional quantitative mathematical models employed in drug development. We propose the elaboration and validation of a logic network for IBD able to capture the information available in the literature that will facilitate the identification/validation of therapeutic targets.

Results

In this article, we propose a logic model for Inflammatory Bowel Disease (IBD) which consists of 43 nodes and 298 qualitative interactions. The model presented is able to describe the pathogenic mechanisms of the disorder and qualitatively describes the characteristic chronic inflammation. A perturbation analysis performed on the IBD network indicates that the model is robust. Also, as described in clinical trials, a simulation of anti-TNFα, anti-IL2 and Granulocyte and Monocyte Apheresis showed a decrease in the Metalloproteinases node (MMPs), which means a decrease in tissue damage. In contrast, as clinical trials have demonstrated, a simulation of anti-IL17 and anti-IFNγ or IL10 overexpression therapy did not show any major change in MMPs expression, as corresponds to a failed therapy. The model proved to be a promising *in silico* tool for the evaluation of potential therapeutic targets, the identification of new IBD biomarkers, the integration of IBD polymorphisms to anticipate responders and non-responders and can be reduced and transformed in quantitative model/s.

the manuscript. Janssen Research and Development provided support in the form of salaries for author AV, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific role of this author is articulated in the 'author contributions' section.

Competing interests: We have the following interests. This study was partly funded by Janssen Research and Development, the employer of An Vermeulen. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

Introduction

Inflammatory bowel disease (IBD) is a complex gastrointestinal tract disorder characterized by a functional impairment of the gut wall affecting patients´ quality of life [1,2]. IBD includes ulcerative colitis (UC) and Crohn's disease (CD). The natural course of IBD is highly variable [3–6] and its etiology is still unknown. The incidence of IBD has dramatically increased worldwide over the past 50 years [7], reaching levels of 24.3 per 100,000 person-years in UC and 20.2 per 100,000 person-years in CD in the developed countries [8].

There is current evidence that Interleukin 6 (IL6), Tumour necrosis factor-alpha (TNFα), Interferon Gamma (IFNɣ), Interleukin 1 beta (IL1ß), Interleukin 22 (IL22), Interleukin 17 (IL17) and Natural Killer cells (NK), among other signalling pathways, play relevant roles in the pathogenesis of IBD, which is a reflection of the complexity of that physiological system [9–12]. That complexity indicates that a universal treatment for IBD may not be feasible for the vast majority of patients [13,14]. In fact, current biological approved treatments are only palliative with a high percentage of non-responders. For example, around 50% of IBD patients treated with the current standard of care, Infliximab (an anti-TNFα) or Vedolizumab (an anti-α4β7 integrin) do not respond satisfactorily to therapy [15,16]. One characteristic of the current IBD biological treatments is that approved therapies target just one signalling pathway, which might explain the high rate of non-responders and the long-term inefficiency of most treatments [15,17]. In addition, there is evidence to suggest that optimal treatment for IBD should involve a combination of different drugs [18,19]. Therefore, there is a need, especially for complex alterations such as immune-mediated diseases, to change the paradigm of drug development, considering the main aspects (targets, cross-talking between pathways, therapy combination) from an integrative and computational perspective.

Given the aforementioned biological complexity of immune-mediated diseases and the fact that current longitudinal data associated with the most relevant elements of the system are scarce, a full parameterization of IBD related systems based on a differential equation model does not yet seem feasible. However, some attempts have been made to describe quantitatively the IBD systems. For example, Wendelsdorf et al., [20] built a quantitative model based on ordinary differential equations. However, some key disease elements, such as cytokines and T cells, were incorporated non-specifically (i.e., all types of cytokine were grouped under the generic element active cytokines) in the model structure, limiting its use to explore potential therapeutic targets. More recently, Dwivendi et al., [21], based on the results of a clinical trial with the anti–IL6R antibody, Tocilizumab, have developed a multiscale systems model in Crohn's disease, limited to the IL6–mediated immune regulation pathway.

Network analysis represents a promising alternative in such data limited circumstances [22–24]. As many molecular pathways in IBD are qualitatively well described, interaction networks may be a suitable approach for characterizing IBD. These networks are simplified representations of biological systems in which the components of the system such as genes, proteins or cells are represented by nodes and the interactions between them by edges [25]. Boolean network models, originally introduced by Kauffman [26,27], represent the simplest discrete dynamic models. These models only assume two discrete states for the nodes of a network, ON or OFF, corresponding to the logic values 1 (active) or 0 (not active, but not necessarily absent) [28]. A well-designed logic model could generate predictive outcomes given a set of initial conditions. Qualitative, logical frameworks have emerged as relevant approaches with different applications, as demonstrated by a growing number of published models [29]. Complementing these applications, several groups have provided various methods and tools to support the definition and analysis of logical models, as it can be seen by the recent achievements of the Consortium for Logical Models and Tools (CoLoMoTo) in logical modelling [30].

There are already several tools for Boolean modeling of regulatory networks in which it is possible to define direct activation-inhibition relationships between the components of the network, such as BoolNet R [31] or GINsim [32]. More recently, the R package SPIDDOR (Systems Pharmacology for effIcient Drug Development On R) among others, has implemented new types of regulatory interactions and perturbations within the system, such as positive and negative modulators and the polymorphism-like alterations, which lead to richer dynamics between the nodes [28].

In the specific case of IBD, there have been initial attempts to develop network models. The multi-state modeling tool published by Mei et al., [33,34] can be considered a proof of concept in the application of these types of networks in mucosal immune responses. However, the number of elements that this model considers and integrates is limited for IBD characterization, since only six different cytokine types are included in the inter-cellular scale.

The objective of the current manuscript is to present a Boolean based network model incorporating the main cellular and protein components known to play a key role in IBD development and progression. The model has been built on well-established experimental knowledge, mostly of human origin, and only including animal data when no other source of information was available. Our aim has been to build a model structure facilitating key aspects in the treatment of immune mediated disease, such as the selection of the most promising combination therapies and the study of the impact of polymorphisms on pathway regulation, thus allowing patient stratification and personalized medicine.

This study provides the scientific community with a (i) computational IBD model implemented in SPIDDOR R package [28], which allows translation of Boolean models (excluding models enclosing temporal operators) to a standard Markup language in Systems Biology for qualitative models (SBML qual [35]) which promotes model interoperability, and (ii) a repository with the main and updated information known of the immune system and IBD, which shows model transparency and allows model reusability. The proposed IBD model can be easily expanded in size and complexity to incorporate new knowledge, or other type of information such as proteomic data. The model presented hereafter is general enough to serve as a skeleton for other relevant immune diseases such as Rheumatoid Arthritis, Psoriasis or Multiple Sclerosis.

The manuscript is organized as follows: In the next section, Results regarding the structure of the model can be graphically visualized, and the ability of the model to recreate certain alterations that have been reported in IBD is demonstrated, as well as the model's capability to reproduce the results from recent clinical trials performed in IBD patients from a high-level perspective. Applications of the model, including its advantages and limitations are then discussed together with ideas for future research. Finally, the Methods section provides a detailed technical description (with the aid of supplementary material) of the network and a description of how simulations, collection, and representation of results have been performed.

Results

Graphical representation, repository, and Boolean functions

The graphical representation of the IBD network is shown in Fig 1. It consists of 43 nodes and 298 qualitative interactions located in three different physiological areas corresponding to (i) the lymph node, (ii) the blood and lymph circulatory system that irrigates the intestinal epithelial cells and (iii) the gut lumen.

Definition of all nodes and the full documented regulatory interactions conforming the model structure can be found in supporting information S1 Table and S2 Table, respectively. The S2 Table is fundamental to understand the rationale for the selection and implementation

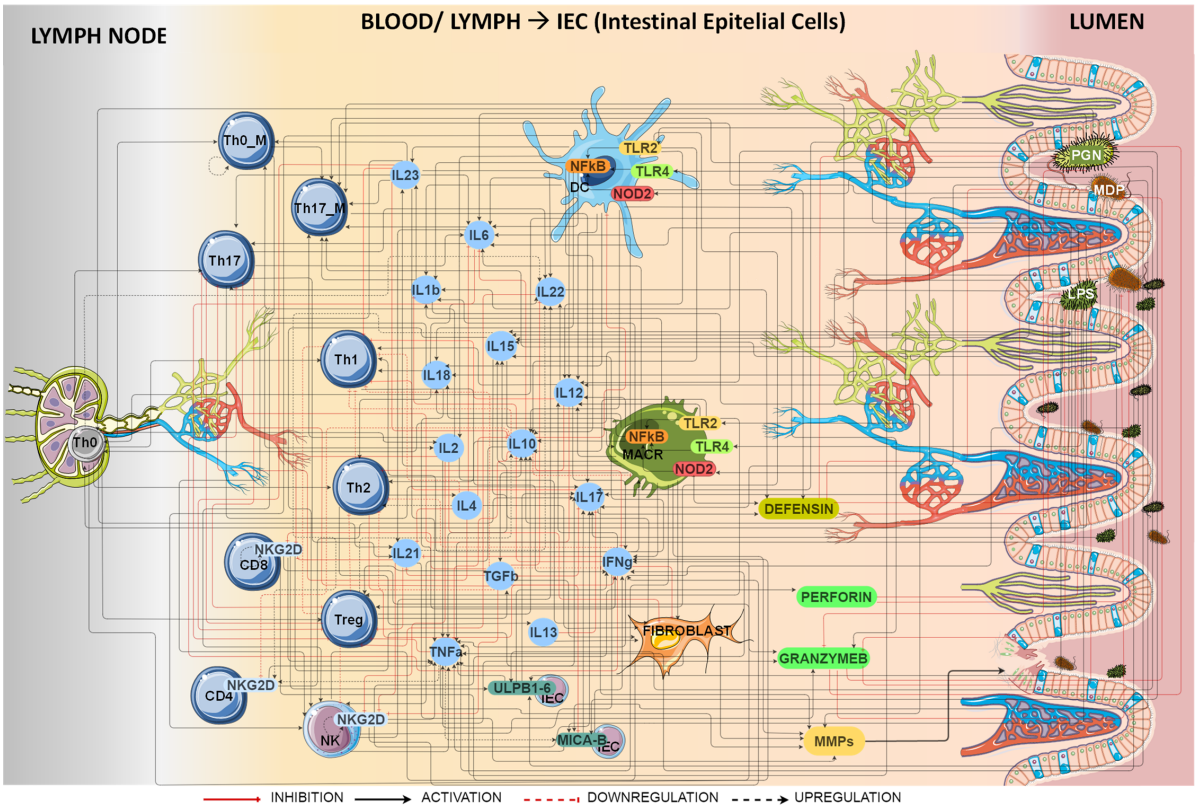


Fig 1. Graphical representation of IBD model. Nodes represent cells, proteins, bacterial antigens, receptors or ligands. Bacterial antigens trigger the IBD immune response through activation of different pattern recognition receptors (TLR2, TLR4 and NOD2) starting the innate and adaptive immune response. Reprinted from [36] under a CC BY license, with permission from the organizers of the 2016 International Conference on Systems Biology, original copyright 2016.

https://doi.org/10.1371/journal.pone.0192949.g001

of the Boolean functions (BF). It was organized to provide a comprehensive summary of the 301 manuscripts (published over the last three decades) used to build the model, highlighting for example whether (i) a specific pathway was reported to be altered in IBD, or (ii) information was supported by human (more than the 80% of the network structure) or animal data.

The Boolean operators used to define the network model of IBD were: the NOT operator which is noted as "!", the AND operator which is noted as "&" and the OR operator which is noted as "|". Recent and innovative modulators and threshold operators previously described by Irurzun-Arana et al., 2017 [28] were also part of the arsenal of Boolean elements used in the model proposed (see S1 File for a detailed description of those additional Boolean elements).

Regarding the input selection, as it is assumed that IBD is caused by intestinal dysbiosis, an environment of different bacteria was recreated selecting three different antigens which are components of most Bacterial Gram positive and Gram negative. Therefore, during the development of the proposed model the following assumptions were made: First, there is a chronic exposure to bacterial antigens: Peptidoglycan (PGN), Lipopolysaccharide (LPS) and Muramyl dipeptide (MDP). PGN is a component of the cell wall of all bacteria, but in particular of gram-positive bacteria, LPS is a component of the outer membrane of Gram-negative bacteria

Table 1. Boolean functions (BF) of the IBD model to simulate the initial conditions.

INITIAL CONDITIONS: CHRONIC EXPOSURE
$PGN = !\,(\bigcap_{i=1}^{AG\_elim=6} PERFOR^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} GRANZB^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} DEF^{t-i})$
$MDP = !\,(\bigcap_{i=1}^{AG\_elim=6} PERFOR^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} GRANZB^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} DEF^{t-i})$
$LPS = !\,(\bigcap_{i=1}^{AG\_elim=6} PERFOR^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} GRANZB^{t-i} \mid \bigcap_{i=1}^{AG\_elim=6} DEF^{t-i})$

https://doi.org/10.1371/journal.pone.0192949.t001

[37], and MDP is a constituent of both Gram-positive and Gram-negative bacteria [38]. All three elicit strong immune responses and seem to play a critical role in the development and pathophysiology of IBD, as it has been hypothesized that the onset or relapse of IBD is triggered by an imbalance in self-microbiota composition than cannot be controlled by immune system [39]. Table 1 lists the initial conditions expressed by the corresponding BF, and shows that the nodes representing antigens are chronically expressed unless the natural antimicrobial peptides perforin (PERFOR), granzyme B (GRANZB) or defensins (DEF) become active.

Second, there is an impairment in antigen elimination in IBD patients [1,40,41], simulated with the threshold operator Ag_elim = 6. The threshold operator means that PERFOR, GRANZB, or DEF inhibit antigen activation when any of these three nodes have been activated for at least 6 consecutive iterations (see Table 1).

Third, the final readout of the network model is the average expression of the output node, Metalloproteinases (MMPs). There is solid evidence that this group of proteins is directly associated with intestinal fibrosis and tissue damage in IBD [42–46] supporting their use as a relevant biomarker in clinical practice as proposed by O'Sullivan et al. [47]. As it can be seen in Table 2, the nodes that directly activate MMPs are the nodes that have relevant roles in the pathogenesis of IBD [9–12,42–44,46,48].

Table 2 contains the full set of BF that modulates the signal initialized by the antigens through the activation of different pattern recognition receptors (TLR2, TLR4 and NOD2 nodes) and the impact on the output node (MMPs) as the recipient of the antigen signal internal modulation. The nodes TNFα or IFNγ have the most complex pathways as can be seen in the corresponding Boolean equations (Table 2).

With the aim of making the network model more accessible to the community it has been uploaded to "The Cell Collective" [49,50] platform (https://www.cellcollective.org/#cb963d7f-75cb-4b2e-8987-0c7592a9c21d). In addition, the supporting information document S2 File provides the network model in text format ready for simulation in the R-based freely available package SPIDDOR [28] and an html tutorial as a guide to reproduce the results (S3 File).

Perturbation analysis and clustering: Network robustness

The results of the network perturbation analysis are presented in Fig 2. The heatmap shows the impact of a single blockage of each node in every network node. The results indicate that most node blockages did not trigger considerable changes, suggesting that the IBD network is robust [51]. Some perturbations led to a higher activation of the nodes, while down regulations were more common. The heatmap was combined with a hierarchical clustering grouping together the nodes that caused similar alterations. Knockout of the NFkß node appeared to be the most relevant alteration as it caused a reduction in expression of many of the nodes that were reported to be overexpressed in IBD patients. The knockout of the Th0 node (representing activated CD4+ T cells) also elicited a reduction in MMPs. The positive effects of the NFkß and Th0 node blockades on MMPs decreased expression, resembled some of the known mechanisms of action of glucocorticoids, inhibitors of T cell activation and proinflammatory

Table 2. Boolean functions (BF) of the IBD model for the internal and the output nodes.

INTERNAL NODES
$TLR2 = PGN$
$TLR4 = LPS$
$NOD2 = MDP$
$NFkB = TLR2 \mid NOD2 \mid TLR4$
$IL6 = (\boldsymbol{MACR}\ \&\ \boldsymbol{PGN}) \mid (DC\ \&\ (LPS \mid PGN)) \mid (Th17\ \&\ IL23) \mid (NFkB\ \&!\ (IL4 \mid IL10))$
$TNFa = ((NFkB\&LPS) \mid (MACR\&(IL2 \mid (IFNg\&LPS) \mid \boldsymbol{PGN})) \mid (NK\&(MDP \mid PGN \mid LPS)\&((IL2 \mid IL12)\&(IL2 \mid IL15)) \mid (FIBROBLAST\&IFNg) \mid (((CD4\_NKG2D \mid CD8\_NKG2D \mid NK\_NKG2D)\&(IEC\_MICA\_B \mid IEC\_ULPB1\_6)))\&!\ (\bigcap_{i=1}^{downreg\_cyt=4} IL10^{t-i}\&(\bigcap_{i=1}^{downreg\_cyt=4} TLR2^{t-i} \mid \bigcap_{i=1}^{downreg\_cyt=4} TLR4^{t-i})\&TNFa)$
$TGFb = (Treg \mid MACR)$
$Th0 = \bigcap_{i=1}^{THR\_Th0=3} LPS^{t-i} \mid \bigcap_{i=1}^{THR\_Th0=3} MDP^{t-i} \mid \bigcap_{i=1}^{THR\_Th0=3} PGN^{t-i}$
$Th0\_M = (Th0\ \&\ (IL23 \mid IL12)) \mid Th0\_M$
$IL18 = ((MACR \mid DC)\&LPS\&NFkB)\&!\ (IL1b\&\bigcup_{i=1}^{downreg\_cyt=4} IL10^{t-i})$
$IFNg = ((NK\&(PGN\mid LPS\mid MDP\mid\&(IL23\mid(IL12\&(IL2\mid IL15\mid IL18)))) \mid (Th0\_M\&(LPS \mid MDP \mid PGN)\ \&(IL12 \mid IL23)) \mid Th1 \mid (((CD8\_NKG2D \mid NK\_NKG2D)\&(IEC\_MICA\_B \mid IEC\_ULPB1\_6)) \mid (Th17\&(PGN \mid LPS\mid MDP)) \mid ((MACR \mid Th0)\&IL18\&IL12))\&!\ ((IFNg\&(\bigcap_{i=1}^{downreg\_cyt=4} TGFb^{t-i} \mid \bigcap_{i=1}^{downreg\_cyt=4} IL10^{t-i} \mid Th2)$
$IL23 = (MACR\ \&\ IL1b) \mid \boldsymbol{DC}$
$IL22 = Th17\mid(NK\&((IL18\&IL12) \mid IL23))\mid CD4\_NKG2D\mid(((IL22\&Th0\&IL21)\&!(\bigcap_{i=1}^{upreg\_cyt=3} IL22^{t-i}\ \&\ \bigcap_{i=1}^{upreg\_cyt=3} Th0^{t-i}\&\bigcap_{i=1}^{upreg\_cyt=3} IL21^{t-i}))\&!\ TGFb)$
$IL21 = \boldsymbol{Th17} \mid ((\boldsymbol{Th0}\ \&\ \boldsymbol{IL6})\ \&!\ (\boldsymbol{IL4} \mid \boldsymbol{IFNg} \mid \boldsymbol{TGFb}))$
$IL17 = (Th17 \mid (Th17\_M\&(LPS \mid MDP \mid PGN)) \mid (CD4\_NKG2D\&(IEC\_MICA\_B \mid IEC\_ULPB1\_6)))\&!\ ((\bigcap_{i=1}^{downreg\_cyt=4} TGFb^{t-i} \mid \bigcap_{i=1}^{downreg\_cyt=4} IL13^{t-i})\&IL17)$
$IL10 = Treg\mid(Th2\ \&!\ IL23)\mid((TLR2\ \&\ NFkB)\ \&!\ (MACR\ \&\ IFNg)) \mid ((MACR\ \&\ LPS)\ \&!\ IL4) \mid (DC\ \&\ LPS)$
$Th17 = ((Th0\&(IL1b \mid IL23 \mid IL6)) \mid ((Th17\&IL23)\&!(\bigcap_{i=1}^{upreg\_cell=2} Th17^{t-i}\&\bigcap_{i=1}^{upreg\_cell=2} Il23^{t-i})))\&!\ (((\bigcup_{i=1}^{downreg\_cell=2} TGFb^{t-i}\mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IL12^{t-i}} \mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IL4^{t-i}} \mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IFNg^{t-i}} \mid \bigcup_{i=1}^{downreg\_cell=2} Treg^{t-i})\&Th17)$
$Th17\_M = ((Th0\_M\ \&\ (PGN \mid MDP \mid LPS))\ \&\ ((IL1b\ \&\ IL6) \mid IL23 \mid IL2)) \mid Th17\_M$
$Th1 = (Th0\&((IL12 \mid IFNg \mid IL18) \mid (DC\&IL12\&IL23\&LPS)))\&!\ ((((\bigcup_{i=1}^{downreg\_cell=2} IL17^{t-i}\ \& \bigcup_{i=1}^{downreg\_cell=2} IL12^{t-i}) \mid (\bigcup_{i=1}^{downreg\_cell=2} Treg^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} Th2^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} TGFb^{t-i} \mid \bigcup_{i=1}^{downreg\_cell=2} IL10^{t-i}\mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IL4^{t-i}}))\&\boldsymbol{Th1})$
$Th2 = (\boldsymbol{Th0}\&(\boldsymbol{IL10} \mid((\boldsymbol{IL18}\&\boldsymbol{IL4})\&!\boldsymbol{IL12}))\mid((\boldsymbol{Th2}\&\boldsymbol{IL4})\&!(\bigcap_{i=1}^{\boldsymbol{upreg\_cell=2}} \boldsymbol{Th2^{t-i}}\&\bigcap_{i=1}^{\boldsymbol{upreg\_cell=2}} \boldsymbol{IL4^{t-i}})))\&!\ ((\bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{Treg^{t-i}}\mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IFNg^{t-i}}\mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{TGFb^{t-i}})\&\boldsymbol{Th2})$
$IL4 = Th2$
$IL15 = (FIBROBLAST\ \&\ (MDP \mid LPS \mid PGN)) \mid (MACR\ \&\ (LPS \mid IFNg))$
$IL12 = ((((MACR \mid DC)\&(LPS \mid PGN)\&IFNg)\&!(IL12\&\bigcup_{i=1}^{downreg\_cyt=4} TNFa^{t-i})) \mid (DC\&IL1b) \mid (IL12\&(IL13 \mid IL4)))\&!((\bigcup_{i=1}^{downreg\_cyt=4} TGFb^{t-i} \mid \bigcup_{i=1}^{downreg\_cyt=4} IL10^{t-i})\&IL12)$
$IL13 = Th2$
$Treg = (\bigcap_{i=1}^{THR\_Th0\_Treg=3} Th0^{t-i}\&(TGFb \mid \boldsymbol{TLR2}))\&!\ ((\bigcup_{i=1}^{downreg\_cell=2} IL6^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} IL21^{t-i}\mid \bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{IL23^{t-i}}\mid \bigcup_{i=1}^{downreg\_cell=2} Th17^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} IL22^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} TNFa^{t-i})\&Treg)$
$NK = (IL15 \mid IL2 \mid IL12 \mid IL23\mid (\boldsymbol{IL18}\&\boldsymbol{IL10}))\&!\ (\bigcup_{i=1}^{\boldsymbol{downreg\_cell=2}} \boldsymbol{Treg^{t-i}}\&\boldsymbol{NK})$
$DEF = IL22 \mid IL17 \mid \bigcap_{i=1}^{THR\_NOD2\_DEF=3} NOD2^{t-i}$
$IL2 = Th0 \mid (Th0\_M\ \&\ (MDP \mid LPS \mid PGN)) \mid DC$
$MACR = (NFkB \mid ((MACR\&(IFNg \mid IL15))\&!\ (\bigcap_{i=1}^{upreg\_cell=2} NFkB^{t-i}\&(\bigcap_{i=1}^{upreg\_cell=2} IFNg^{t-i} \mid \bigcap_{i=1}^{upreg\_cell=2} IL15^{t-i}))))\&!\ (\bigcup_{i=1}^{downreg\_cell=2} IL10^{t-i}\&MACR)$
$DC = NFkB\&!\ (\bigcup_{i=1}^{downreg\_cell=2} IL10^{t-i}\&DC)$
$IEC\_MICA\_B = ((LPS \mid MDP \mid PGN) \mid (IEC\_MICA\_B\&TNFa)\&!\ (\bigcap_{i=1}^{upreg\_rec=2} IEC\_MICA\_B^{t-i}\& \bigcap_{i=1}^{upreg\_rec=2} TNFa^{t-i}))\&!\ TGFb$
$IEC\_ULPB1\_6 = CD8\_NKG2D\ \&\ (LPS\mid MDP\mid PGN)$
$CD8\_NKG2D = (LPS \mid PGN \mid MDP)\&!((\bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_MICA\_B^{t-i} \mid \bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_ULPB\_1\_6^{t-i} \mid (\bigcup_{i=1}^{downreg\_cell=2} IL21^{t-i}\&\bigcup_{i=1}^{downreg\_cell=2} IL2^{t-i}))\&CD8\_NKG2D)$
$NK\_NKG2D = (LPS\mid PGN\mid MDP)\&!\ (\bigcup_{i=1}^{downreg\_cell=2} TGFb^{t-i}\mid \bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_MICA\_B^{t-i} \mid \bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_ULPB\_1\_6^{t-i} \mid (\bigcup_{i=1}^{downreg\_cell=2} IL21^{t-i}\&\bigcup_{i=1}^{downreg\_cell=2} IL12^{t-i}))\&NK\_NKG2D)$
$CD4\_NKG2D = (LPS \mid PGN \mid MDP \mid (CD4\_NKG2D\&(IL15 \mid TNFa))\&!\ ((\bigcap_{i=1}^{upreg\_rec=2} CD4\_NKG2D^{t-i}\ \&(\bigcap_{i=1}^{upreg\_rec=2} IL15^{t-i}\mid \bigcap_{i=1}^{upreg\_rec=2} TNFa^{t-i}))\&!\ ((\bigcup_{i=1}^{downreg\_cell=2} IL10^{t-i} \mid \bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_MICA\_B^{t-i} \mid \bigcap_{i=1}^{THR\_LIGANDS\_NKG2D=3} IEC\_ULPB\_1\_6^{t-i})\&CD4\_NKG2D))$
$FIBROBLAST = ((MACR\&(IL4 \mid IL13 \mid TGFb))\mid IL2)\&!((\bigcup_{i=1}^{downreg\_cell=2} IFNg^{t-i}\mid \bigcup_{i=1}^{downreg\_cell=2} IL12^{t-i})\ \&FIBROBLAST)$
$PERFOR = NK \mid NK\_NKG2D$

(*Continued*)

Table 2. (Continued)

GRANZB = CD8_NKG2D | NK | NK_NKG2D | (DC &! (LPS | PGN))
OUTPUT NODE
MMPs = (MACR & TNFa) | (FIBROBLAST & (IL21 | IL17 | IL1b | TNFa))

Bold text within Boolean equations indicates that the information belongs to animal data

https://doi.org/10.1371/journal.pone.0192949.t002

cytokines, as well as potent suppressors of the effector function of monocyte-macrophage and fibroblastic activity, interfering with the NFκB inflammatory signal [52–54].

Network accuracy and validation

Experimental and clinical information. Simulations of chronic infection in IBD individuals show that the model reproduced satisfactorily experimental and clinical information (summarized in Table 3 and supporting information S3 Table). Fig 3 shows the results of the simulation for each network node after reaching the attractor state for virtual healthy and IBD subjects. In total, 31 upregulations in experimental studies were replicated with our simulations. Similarly, the 9 nodes reported as altered appeared upregulated in the simulations, and finally, the three nodes whose profiles were not known also proved to be upregulated.

Clinical trials. In our simulations, three drugs that have failed to prove clinical efficacy in clinical trials (anti-IL17, anti-IFNγ and rhuIL-10) also exhibited no benefit in the simulated surrogate for the disease score (Fig 4). Simulations with anti-TNFα, a biologic therapy approved for IBD, showed a decrease in the disease score. Simulations with anti-IL12-IL23, a recently approved therapy for IBD, showed a slight decrease in MMPs and anti-IL2 therapy simulation showed a decrease similar to anti-TNFα. In addition, the new promising therapy (GMA), equivalent to an anti-MACR in our model showed a decrease in MMPs similar to that for anti-TNFα.

Discussion

In the current study, we present a Systems Pharmacology (SP) network model for IBD based on the main cells and proteins involved in the disease. Our analysis appears timely, as IBD has recently been attracting increasing attention [55–59]. We attempted to meet one of the major challenges in inflammatory bowel disease (IBD) which is the integration of IBD-related information to construct a predictive model. We are not the only ones following this line of research, as Lauren A Peters et al. have very recently performed a key driver analysis to identify the genes predicted to modulate network regulatory states associated with IBD [55]. Both analyses could be integrated in the future and inform our post-transcriptomic network with the key driver genes identified by Lauren A Peters et al. [55].

In comparison with the previous quantitative approaches for IBD [20,21,33,34], our model identified Naive CD4+ T Cells, Macrophages and Fibroblasts cells as relevant in IBD. Also, in addition to the six interleukins (TGFß, IL6, IL17, IL10, IL12 and IFNγ) considered by Mei et al. [33,34] our network involves 10 interleukins more which could represent possible IBD biomarkers [60]. The procedure to evaluate the potential role of the different components on the disease as plausible biomarkers, would be equal to the one described in section 4.5 (perturbation analysis and clustering), focussing on the changes in the output node.

In the validation of network models, robustness and practical applicability represent critical aspects. The fact that the information gathered from the literature was obtained under very

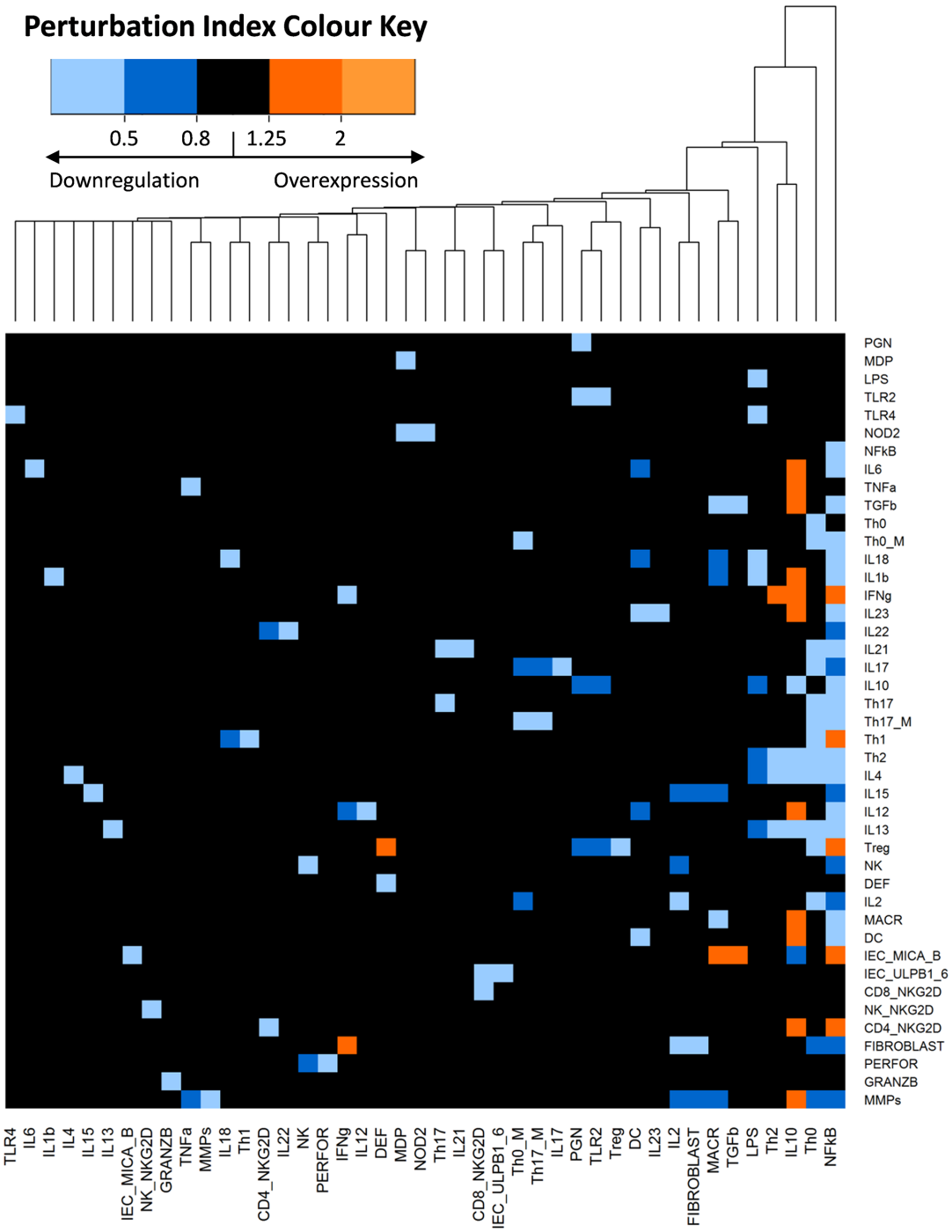

Perturbation Index Colour Key
0.5
0.8
1.25
2
Downregulation
Overexpression
PGN
MDP
LPS
TLR2
TLR4
NOD2
NFkB
IL6
TNFa
TGFb
Th0
Th0_M
IL18
IL1b
IFNg
IL23
IL22
IL21
IL17
IL10
Th17
Th17_M
Th1
Th2
IL4
IL15
IL12
IL13
Treg
NK
DEF
IL2
MACR
DC
IEC_MICA_B
IEC_ULPB1_6
CD8_NKG2D
NK_NKG2D
CD4_NKG2D
FIBROBLAST
PERFOR
GRANZB
MMPs
TLR4
IL6
IL1b
IL4
IL15
IL13
IEC_MICA_B
NK_NKG2D
GRANZB
TNFa
MMPs
IL18
Th1
CD4_NKG2D
IL22
NK
PERFOR
IFNg
IL12
DEF
MDP
NOD2
Th17
IL21
CD8_NKG2D
IEC_ULPB1_6
Th0_M
Th17_M
IL17
PGN
TLR2
Treg
DC
IL23
IL2
FIBROBLAST
MACR
TGFb
LPS
Th2
IL10
Th0
NFkB

Fig 2. IBD network perturbation analysis and clustering. The heatmap indicates the effect of single blockage of each node (columns) in every network node (rows). The colour in each cell corresponds to the Perturbation Index (PI) of the nodes. When there is no change in the expression of the node, the cells of the heatmap would be black, having a value between 0.8 and 1.25 in their PIs. Otherwise, when the perturbation causes an overexpression in a node, the cell in the heatmap would be orange coloured, with PIs values greater than 1.25. On the contrary, a value of 0.8 or smaller, blue colour, indicates that the perturbation causes a downregulation of the node. The numeric scale in the legend represents different values of the nodes PI under different perturbations. Nodes that induce similar alterations are hierarchically clustered.

https://doi.org/10.1371/journal.pone.0192949.g002

different experimental designs/conditions/methodologies, represents a challenge with respect to validation. This led us to propose and adopt a novel strategy consisting of the comparison of the results of model-based virtual pathway simulations with those reported in the literature for IBD patients. Using this approach, we obtained a qualitative reproduction of IBD in which all the network elements that have been reported as upregulated in IBD patients appeared upregulated in our simulation results. The perturbation analysis of the network was performed by a single blockage in each node to analyse how that type of alteration propagates through the entire network reflecting the case of single polymorphisms, which represents the simplest case of IBD disease. Despite of the simplicity of this analysis, the results obtained from the model accuracy and validation procedures are encouraging. Results from the perturbation analysis indicate that the proposed network model is robust, as alteration in most nodes did not trigger considerable changes in the network [61].

Once validated and checked for robustness, the network was challenged to qualitatively reproduce the readouts of five different therapies reported in experimental and clinical studies. The outcome of this challenge was similar to the clinical output in IBD patients. By the simulation of TNFα or MACR knockout (simulating Granulocyte and Monocyte Apheresis), a decrease in MMPs node was observed, which is in line with therapy success in clinical practice by a decrease in Crohn's Disease Activity Index (CDAI) Score [42–46],[62–68]. On other hand, IL17 or IFNγ knockout or IL10 overexpression did not show major change in MMPs expression, suggested a failed therapy as was indeed found in clinical practice [69–72].

Surprisingly, the model shows that a knockout of IL2 leads to a reduction in MMPs similar to that of a knockout of TNFα, even when previous results of clinical trials with Basiliximab or Daclizumab (monoclonal antibodies that bind to the interleukin 2 receptor CD25) in Ulcerative Colitis have failed to show superiority to corticosteroids alone [73,74]. The mechanism of

Table 3. Expression of network nodes in IBD patients.

NODE	EXPRESSION	NODE	EXPRESSION	NODE	EXPRESSION	NODE	EXPRESSION
PGN MDP LPS	Altered	IL1b	Upregulated	Th2	Upregulated	DC	Downregulated in Blood-Upregulated in mucosa
TLR2	Upregulated	IFNg	Upregulated	IL4	Altered	IEC_MICA_B	Upregulated
TLR4	Upregulated	IL23	Upregulated	IL15	Upregulated	IEC_ULPB1_6	Upregulated
NOD2	Altered	IL22	Upregulated	IL12	Upregulated	CD8_NKG2D	Upregulated
NFkB	Altered	IL21	Upregulated	IL13	Upregulated	NK_NKG2D	Unknown
IL6 TNFa	Upregulated Upregulated	IL17	Upregulated	Treg	Downregulated in Blood-Upregulated in mucosa	CD4_NKG2D	Upregulated
TGFb	Upregulated	IL10	Upregulated	NK	Upregulated	FIBROBLAST	Upregulated
Th0	Unknown	Th17	Upregulated	DEF	Altered	MMPs	Upregulated
Th0_M	Upregulated	Th17_M	Upregulated	IL2	Upregulated	PERFOR	Altered
IL18	Upregulated	Th1	Altered	MACR	Unknown	GRANZB	Upregulated

A total of 31 nodes are reported as upregulated in IBD patients, 9 are reported to be altered (when different reports from literature are inconclusive or contradictory) and 3 nodes are unknown.

https://doi.org/10.1371/journal.pone.0192949.t003

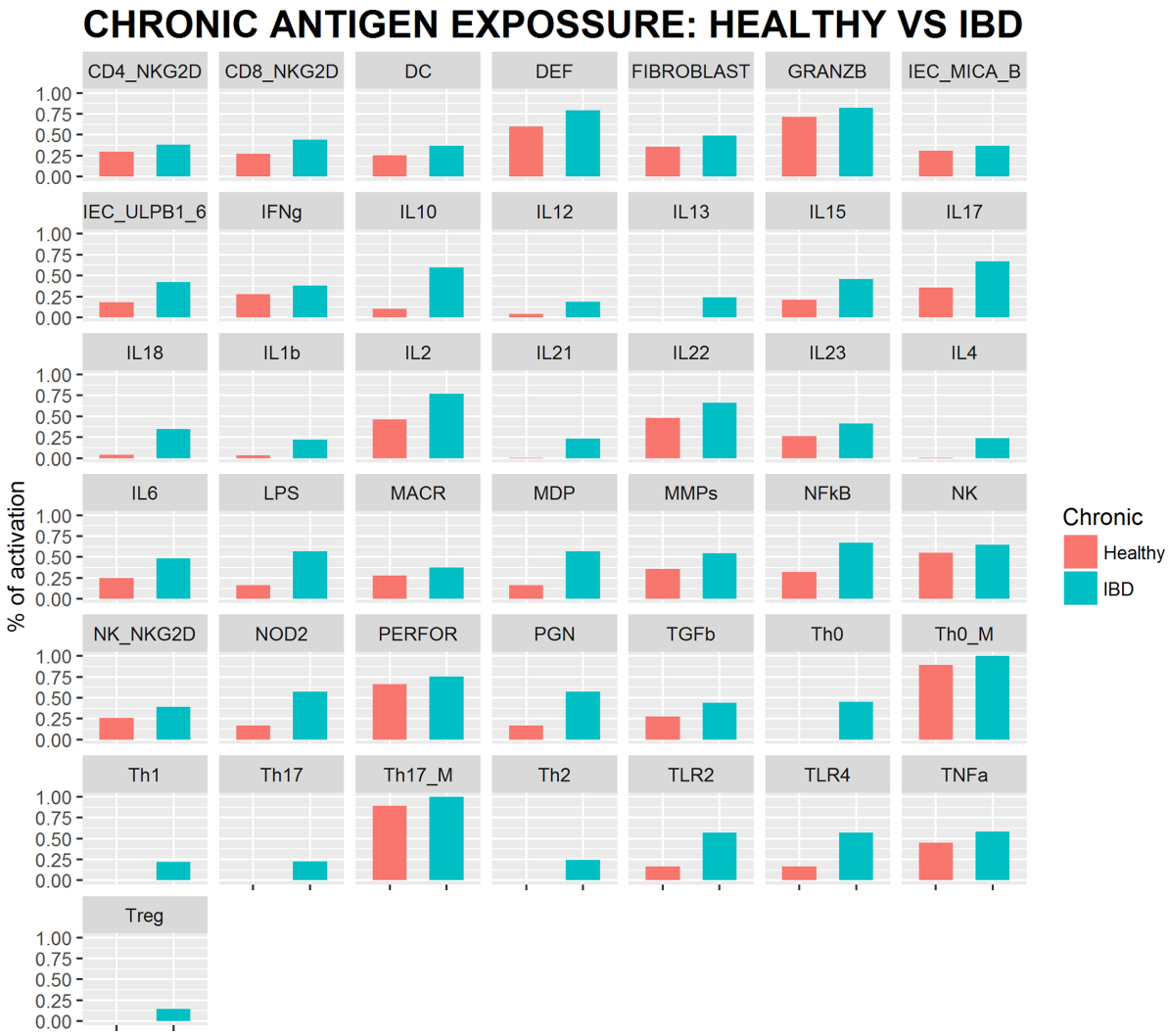


Fig 3. IBD network simulation results. Attractor state of every network node for healthy and IBD simulated individuals under chronic antigen exposure.

https://doi.org/10.1371/journal.pone.0192949.g003

action of corticosteroids has not been fully described, yet it is known that corticosteroids cause diminished levels of IL2 mRNA [75,76]. Together with the rest of corticosteroid inhibitory mechanisms, this would be the reason why Basiliximab or Daclizumab do not show superiority to corticosteroids alone.

Among the potential applications the current network supports: (i) biomarker selection given that the cytokines TNFα, IL21, IL17 and IL1ß, which can be easily measured in peripheral plasma with different Enzyme-linked immunosorbent assay (ELISA) kits [77,78], are the model components directly related to MMPs activation, (ii) search for optimal combination therapy to overcome the high attrition rates in phase clinical trials with single therapies which

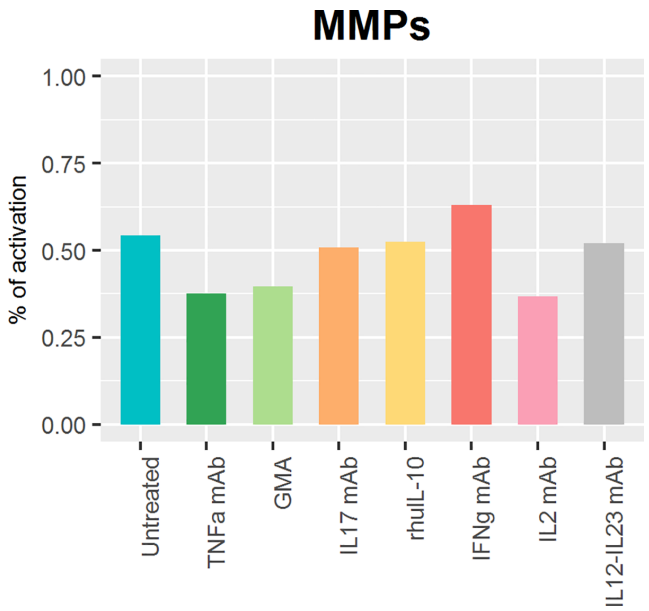


Fig 4. Comparison of MMPs expression after the simulation in IBD simulated individuals of different therapies. Simulated therapies: Anti-TNFα, GMA therapy (equivalent of knock out our MACR node), anti-IL17, human recombinant IL10 (rhulL-10), anti-IFNγ, anti-IL2 and anti-IL12-IL23. Comparing with untreated simulation, we can see a 30.7%, a 27.1%, a 31.9% and a 4.1% decrease in the MMPs expression simulating anti-TNFα, GMA therapy, anti-IL2 and anti-IL12-IL23 respectively. There is no major change in MMPs expression for the two which failed in clinical trials anti-IL17 (a 6.5% decrease) and human recombinant IL10 (a 3.2% decrease). Otherwise, anti-IFNγ therapy simulation shows an increase in MMPs expression of 16.0% compared to Untreated.

https://doi.org/10.1371/journal.pone.0192949.g004

are due mainly to lack of efficacy [79], and (iii) management of multiscale information such as the integration of proteomic gene expression data [55] accounting for IBD polymorphisms to anticipate responders and non-responders. With such a type of data able to correlate a genetic alteration with a decrease or an increase in protein expression, it would be possible to simulate specific genetic alteration by altering the protein expression. This would allow one of the limitations of the current network at the present time to be overcome with regard to the effects of Ustekinumab, a monoclonal antibody targeting free IL12 and IL23, which has been recently approved for moderately to severely active Crohn's disease in adults who have failed to treatment with immunomodulators, or more than one TNFα blocker [80]. Simulation results based on the known mechanisms of Ustekinumab showed just a 4.1% decrease in tissue damage. On the other hand, when simulating TNFα blocker effects, tissue damage decreased by 30.6% even though a substantial percentage of patients showed poor control of the disease after treatment with anti-TNFα antibody [15,16].

We emphasize that the proposed network model is fully accessible which allows it to undergo immediate testing and further development. In that respect it should be noted that although our model intended to include information of human origin exclusively, some critical pathways had to be complemented with animal derived data (although in the current case the percentage of human supported pathways is greater than in previous computational models [20,81,82]), but we are aware of the wide differences in the immune system between species [83–85].

This study addresses the goals of systems pharmacology by effectively encompassing prior knowledge to generate a mechanistic and predictive understanding at the systems level for IBD. Semi-quantitative understanding at the network level is necessary prior to the generation of detailed quantitative models for within-host disease dynamics. The current IBD model and the companion literature summary archive will drive the development of a dynamic (i.e., ordinary differential equation driven) model involving meaningful parameters capable of simulating longitudinal data, and allowing model reduction as well the goal of parameter estimation during the clinical stages of the drug development process. In addition, our IBD network can be extended to other inflammatory diseases, as main pathways in the model are common to most inflammatory conditions [86,87], and the outputs of our nodes could also serve as inputs to broader-scale logic models; for example, incorporating structures from available logic models of some of our nodes such as fibroblast [61], IL1b or IL6 [88].

In summary, we present a network model for inflammatory bowel disease which is available and ready to be used and can cope with (multi-scale) model extensions. It is supported by a comprehensive repository summarizing the results of the most relevant literature in the field. This model proved to be promising for the *in silico* evaluation of potential therapeutic targets, the search for pathway specific biomarkers, the integration of polymorphisms for patient stratification, and can be reduced and transformed in quantitative model/s.

Methods

Literature search and data selection

The network model is based on an exhaustive bibliographic review focusing on the essential components of IBD, as previously performed by Ruiz-Cerdá et al., in their systems pharmacology approach for lupus erythematosus [23]. Our review included around 620 papers published between October 1984 and September 2017, yet the most common reviewed articles were from 2007 or later (76%). The search of the relevant literature was made through Medical Subject Headings (MeSH) terms using different search engines such as PubMed, clinicaltrials.gov or google scholar. MeSH terms were focused on the combination of keywords and free words including: (i) relevant network components (ej."IL6") involved in the pathogenesis of IBD, (ii) nodes that have been reported to be altered in IBD (ej. "IL6 AND IBD") and (iii) nodes directly affecting the expression of the nodes selected in (i) and (ii) (ej. "DC AND IL6"). The internal nodes selection was made according to the reported upregulated components in IBD patients together with the nodes (immune system cells) which are necessary to link the upregulated nodes, which were established as internal nodes. Only original papers with a clear description of experimental conditions were considered to identify the relationships between the components of the biological network. Due to the reported differences between animal and human immunology [83–85], in only few cases were animal data considered to connect nodes of critical pathways when no human data were available.

Annotation and system representation

Annotation was crucial to organize the available literature according to its relevance. S2 Table from supplementary information shows the way the information was organized for building the network. S2 Table includes every node definition and the relationships between the nodes. Annotation included the identification of the main elements (antigens, cytokines, cells, proteins, membrane receptors and ligands) of IBD disease.

The IBD model will be freely accessible to the public through the "The Cell Collective" repository https://cellcollective.org/#cb963d7f-75cb-4b2e-8987-0c7592a9c21d.

Boolean network building and r implementation

The collection of qualitative relationships extracted from the literature was transformed into a logical model as described before by Ruiz-Cerdá et al. [23]. Logic networks capture the dynamics of their components, called nodes, after selected stimuli or initial conditions [89,90]https://paperpile.com/c/XvtklO/p0BRz+YiQ4q. In these models the relationships of activation or inhibition between nodes are described as combinations of the logic operators: AND, OR and NOT condensed in a mathematical expression called a Boolean function for each node. Positive and negative modulators, and thresholds as previously described by Ruiz-Cerdá et al.[23] and Irurzun-Arana et al. [28] were also considered to resemble better the biological system. Boolean network building and R implementation from S1 File gives a more detailed explanation of the modulators used in the model.

Simulations

The set of combined Boolean functions for the IBD model was implemented SPIDDOR [28], using RStudio Version 0.99.442. Simulations with 25 repetitions over 5000 iterations were performed. According to preliminary experiments, these simulation conditions were required to achieve the steady state of the network called attractor [91–93]. An attractor can be a fixed-point if it composed of one state, a simple cycle if consists of more than one state that oscillates in a cycle or a complex attractor if a set of steady-states oscillate irregularly. In each simulation, a node can show two possible values in each iteration: 0 (deactivated) or 1 (activated). The percentage of activation of the output node (MMPs) calculated at the attractor state was used as the readout summary of the simulation exercises, as this group of proteins are directly associated with intestinal fibrosis and tissue damage in IBD [42–46].

Each node was updated asynchronously [94–96] according to its Boolean function that defines the dynamics of the system. Initial conditions are explained in detail in "Simulations" from S1 File.

Perturbation analysis and clustering

Robustness can be defined as the system's ability to function normally under stochastic perturbations [96]. The investigation of robustness in Boolean networks generally focuses on the dependence between robustness and network connectivity [97]. We performed a perturbation analysis in our IBD model to study robustness by simulating the effect of the single blockage of each node on every other node of the network [51]. This simulation was performed by using the *KO_matrix.f* function from SPIDDOR package with 24 repetitions over 999 iterations under asynchronous updating.

Results from the simulations described above were represented as heatmaps with dendrograms in which the number of rows and columns is equal to the number of nodes in the network (Fig 2). The colour in each cell of the heatmap corresponds to the Perturbation Index(PI) of the nodes, which is the probability ratio between the perturbed and the normal conditions as described by Irurzun-Arana et al. [28]. A hierarchical clustering method [98] was applied to further study which nodes cause similar alterations in the system.

Network accuracy and validation

Accuracy was evaluated comparing the alterations reported in the literature for IBD patients with the simulations of chronic antigen exposure for IBD or healthy individuals.

A literature search of every node expression in IBD patients was performed, and the gathered information is condensed in S3 Table including three categories: up-, down-regulated, or

altered, whether the levels in CD, UC or both (IBD) with respect to healthy volunteers are higher, lower, or inconclusive and/or contradictory, respectively.

For validation purposes, model simulations were compared against available results from clinical trials performed in IBD, CD or UC until the beginning of 2017 in https://www.clinicaltrials.gov/. All the molecules tested in clinical trials, whose mechanism of action is known and whose target were included in our network, were tested with the model. The network was evaluated comparing simulations and reported outcomes from clinical trials for six investigated molecules: anti-TNFα [62–65] and anti-IL12-IL23 [80], two monoclonal antibodies (mAb) approved for IBD disease, anti-IFNγ [69,70], anti-IL17 [72], anti-IL2 [73,74] and human recombinant IL10 (rhuIL-10) [71] which failed in clinical trials. Also a new promising therapy: Granulocyte and Monocyte Apheresis (GMA) [66–68] was tested. The reported CDAI (Crohn Disease Activity Index) was compared with the average expression of the MMPs output node in the attractor state.

Supporting information

S1 Table. Abbreviations. List of abbreviations.
(PDF)

S2 Table. IBD Network Repository. Table of nodes and interactions supported by references.
(PDF)

S3 Table. IBD_validation. Table of alterations in patients of IBD network nodes supported by references.
(PDF)

S1 File. Supporting_Information_Methods. Document with detailed description of the methodology.
(DOCX)

S2 File. IBD.txt. Text document with the Boolean functions written in SPIDDOR nomenclature for iBD simulation.
(TXT)

S3 File. User_Guide_SPIDDOR_IBD.html. Html tutorial about how to reproduce the results from the present manuscript with the SPIDDOR package.
(HTML)

Acknowledgments

We would like to thank The Cell Collective team, specially to Tomas Helikar, for their help in building the model in their platform and making it more accessible to the community.

Author Contributions

Conceptualization: Violeta Balbas-Martinez, José David Gómez-Mantilla, Iñaki F. Trocóniz.

Data curation: Violeta Balbas-Martinez, Leire Ruiz-Cerdá, Ignacio González-García.

Formal analysis: Violeta Balbas-Martinez, Leire Ruiz-Cerdá, Itziar Irurzun-Arana, José David Gómez-Mantilla.

Funding acquisition: An Vermeulen, Iñaki F. Trocóniz.

Investigation: Violeta Balbas-Martinez, Leire Ruiz-Cerdá, Ignacio González-García, José David Gómez-Mantilla.

Methodology: Violeta Balbas-Martinez, Itziar Irurzun-Arana, Ignacio González-García, José David Gómez-Mantilla, Iñaki F. Trocóniz.

Project administration: An Vermeulen, Iñaki F. Trocóniz.

Software: Violeta Balbas-Martinez, Itziar Irurzun-Arana, Ignacio González-García, José David Gómez-Mantilla.

Supervision: An Vermeulen, José David Gómez-Mantilla, Iñaki F. Trocóniz.

Validation: Violeta Balbas-Martinez, José David Gómez-Mantilla.

Visualization: Violeta Balbas-Martinez.

Writing – original draft: Violeta Balbas-Martinez, Iñaki F. Trocóniz.

Writing – review & editing: Violeta Balbas-Martinez, José David Gómez-Mantilla, Iñaki F. Trocóniz.

References

1. Wehkamp J, Götz M, Herrlinger K, Steurer W, Stange EF. Inflammatory Bowel Disease. Dtsch Arztebl Int. 2016; 113: 72–82. https://doi.org/10.3238/arztebl.2016.0072 PMID: 26900160
2. Matricon J, Barnich N, Ardid D. Immunopathogenesis of inflammatory bowel disease. Self Nonself. 2010; 1: 299–309. https://doi.org/10.4161/self.1.4.13560 PMID: 21487504
3. Solberg IC, Lygren I, Jahnsen J, Aadland E, Høie O, Cvancarova M, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). Scand J Gastroenterol. 2009; 44: 431–440. https://doi.org/10.1080/00365520802600961 PMID: 19101844
4. Faubion WA Jr, Loftus EV Jr, Harmsen WS, Zinsmeister AR, Sandborn WJ. The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology. 2001; 121: 255–260. PMID: 11487534
5. Loftus EV, Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. Aliment Pharmacol Ther. Wiley Online Library; 2002; 16: 51–60. PMID: 11856078
6. Henriksen M, Jahnsen J, Lygren I, Sauar J, Kjellevold Ø, Schulz T, et al. Ulcerative colitis and clinical course: results of a 5-year population-based follow-up study (the IBSEN study). Inflamm Bowel Dis. 2006; 12: 543–550. https://doi.org/10.1097/01.MIB.0000225339.91484.fc PMID: 16804390
7. Gasparetto M, Guariso G. Highlights in IBD Epidemiology and Its Natural History in the Paediatric Age. Gastroenterol Res Pract. 2013; 2013: 829040. https://doi.org/10.1155/2013/829040 PMID: 24454343
8. Molodecky NA, Soon S, Rabi DM, Ghali WA, Ferris M. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology. Elsevier; 2012; Available: http://www.sciencedirect.com/science/article/pii/S0016508511013783
9. Brand S, Beigel F, Olszak T, Zitzmann K, Eichhorst ST, Otte J-M, et al. IL-22 is increased in active Crohn's disease and promotes proinflammatory gene expression and intestinal epithelial cell migration. Am J Physiol Gastrointest Liver Physiol. 2006; 290: G827–38. https://doi.org/10.1152/ajpgi.00513.2005 PMID: 16537974
10. Yadav PK, Chen C, Liu Z. Potential role of NK cells in the pathogenesis of inflammatory bowel disease. J Biomed Biotechnol. 2011; 2011: 348530. https://doi.org/10.1155/2011/348530 PMID: 21687547
11. Rovedatti L, Kudo T, Biancheri P, Sarra M, Knowles CH, Rampton DS, et al. Differential regulation of interleukin 17 and interferon gamma production in inflammatory bowel disease. Gut. 2009; 58: 1629–1636. https://doi.org/10.1136/gut.2009.182170 PMID: 19740775
12. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. Nat Clin Pract Gastroenterol Hepatol. 2006; 3: 390–407. https://doi.org/10.1038/ncpgasthep0528 PMID: 16819502
13. Hamedani R, Feldman RD, Feagan BG. Review article: Drug development in inflammatory bowel disease: budesonide—a model of targeted therapy. Aliment Pharmacol Ther. 1997; 11 Suppl 3: 98–107; discussion 107–8.

14. de Lange KM, Barrett JC. Understanding inflammatory bowel disease via immunogenetics. J Autoimmun. 2015; https://doi.org/10.1016/j.jaut.2015.07.013 PMID: 26257098

15. Naija N, Karoui S, Serghini M, Kallel L, Boubaker J, Filali A. [Management of failure of infliximab in inflammatory bowel disease]. Tunis Med. 2011; 89: 517–521. PMID: 21681712

16. Shelton E, Allegretti JR, Stevens B, Lucci M, Khalili H, Nguyen DD, et al. Efficacy of Vedolizumab as Induction Therapy in Refractory IBD Patients: A Multicenter Cohort. Inflamm Bowel Dis. 2015; 21: 2879–2885. https://doi.org/10.1097/MIB.0000000000000561 PMID: 26288002

17. Raine T. Vedolizumab for inflammatory bowel disease: Changing the game, or more of the same? United European Gastroenterol J. 2014; 2: 333–344. https://doi.org/10.1177/2050640614550672 PMID: 25360311

18. Coskun M, Vermeire S, Nielsen OH. Novel Targeted Therapies for Inflammatory Bowel Disease. Trends Pharmacol Sci. 2016; https://doi.org/10.1016/j.tips.2016.10.014 PMID: 27916280

19. Grevenitis P, Thomas A, Lodhia N. Medical Therapy for Inflammatory Bowel Disease. Surg Clin North Am. 2015; 95: 1159–82, vi. https://doi.org/10.1016/j.suc.2015.08.004 PMID: 26596920

20. Wendelsdorf K, Bassaganya-Riera J, Hontecillas R, Eubank S. Model of colonic inflammation: immune modulatory mechanisms in inflammatory bowel disease. J Theor Biol. 2010; 264: 1225–1239. https://doi.org/10.1016/j.jtbi.2010.03.027 PMID: 20362587

21. Dwivedi G, Fitz L, Hegen M, Martin SW, Harrold J, Heatherington A, et al. A Multiscale Model of Interleukin-6–Mediated Immune Regulation in Crohn's Disease and Its Application in Drug Discovery and Development. CPT: Pharmacometrics & Systems Pharmacology. 2014; 3: 1–9.

22. Chen S, Jiang H, Cao Y, Wang Y, Hu Z, Zhu Z, et al. Drug target identification using network analysis: Taking active components in Sini decoction as an example. Sci Rep. 2016; 6: 24245. https://doi.org/10.1038/srep24245 PMID: 27095146

23. Ruiz-Cerdá ML, Irurzun-Arana I, González-Garcia I, Hu C, Zhou H, Vermeulen A, et al. Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. Eur J Pharm Sci. 2016; https://doi.org/10.1016/j.ejps.2016.04.010 PMID: 27080094

24. Zhou W, Wang Y, Lu A, Zhang G. Systems Pharmacology in Small Molecular Drug Discovery. Int J Mol Sci. 2016; 17: 246. https://doi.org/10.3390/ijms17020246 PMID: 26901192

25. Le Novère N. Quantitative and logic modelling of molecular and gene networks. Nat Rev Genet. 2015; 16: 146–158. https://doi.org/10.1038/nrg3885 PMID: 25645874

26. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969; 22: 437–467. PMID: 5803332

27. Fox RF. Review of Stuart Kauffman, The Origins of Order: Self-Organization and Selection in Evolution. Biophys J. 1993; 65: 2698–2699.

28. Irurzun-Arana I, Pastor JM, Trocóniz IF, Gómez-Mantilla JD. Advanced Boolean modeling of biological networks applied to systems pharmacology. Bioinformatics. 2017; https://doi.org/10.1093/bioinformatics/btw747 PMID: 28073755

29. Naldi A, Monteiro PT, Müssel C, Consortium for Logical Models and Tools, Kestler HA, Thieffry D, et al. Cooperative development of logical modelling standards and tools with CoLoMoTo. Bioinformatics. 2015; 31: 1154–1159. https://doi.org/10.1093/bioinformatics/btv013 PMID: 25619997

30. Abou-Jaoudé W, Traynard P, Monteiro PT, Saez-Rodriguez J, Helikar T, Thieffry D, et al. Logical Modeling and Dynamical Analysis of Cellular Networks. Front Genet. 2016; 7: 94. https://doi.org/10.3389/fgene.2016.00094 PMID: 27303434

31. Müssel C, Hopfensitz M, Kestler HA. BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. Bioinformatics. Oxford University Press; 2010; 26: 1378–1380. https://doi.org/10.1093/bioinformatics/btq124 PMID: 20378558

32. Chaouiya C, Naldi A, Thieffry D. Logical Modelling of Gene Regulatory Networks with GINsim. In: van Helden J, Toussaint A, Thieffry D, editors. Bacterial Molecular Networks: Methods and Protocols. New York, NY: Springer New York; 2012. pp. 463–479.

33. Mei Y, Abedi V, Carbo A, Zhang X, Lu P, Philipson C, et al. Multiscale modeling of mucosal immune responses. BMC Bioinformatics. 2015; 16 Suppl 12: S2.

34. Bassaganya-Riera J. Computational Immunology: Models and Tools. Academic Press; 2015.

35. Chaouiya C, Bérenguier D, Keating SM, Naldi A, van Iersel MP, Rodriguez N, et al. SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. BMC Syst Biol. 2013; 7: 135. https://doi.org/10.1186/1752-0509-7-135 PMID: 24321545

36. Violeta Balbas-Martinez, Leire Ruiz-Cerdá, Itziar Irurzun-Arana, Ignacio González-García, José David Gómez-Mantilla and Iñaki F. Trocóniz. Systems Pharmacology model for Inflammatory Bowel Disease

(IBD). In: International Conference on Systems Biology 2016, editor. https://doi.org/10.3252/pso.eu.17ICSB.2016

37. Smits HH, van Beelen AJ, Hessle C, Westland R, de Jong E, Soeteman E, et al. Commensal Gram-negative bacteria prime human dendritic cells for enhanced IL-23 and IL-27 expression and enhanced Th1 development. Eur J Immunol. 2004; 34: 1371–1380. https://doi.org/10.1002/eji.200324815 PMID: 15114670

38. Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J, et al. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem. 2003; 278: 5509–5512. https://doi.org/10.1074/jbc.C200673200 PMID: 12514169

39. Wehkamp J, Frick J-S. Microbiome and chronic inflammatory bowel diseases. J Mol Med. 2016; https://doi.org/10.1007/s00109-016-1495-z PMID: 27988792

40. Buttó LF, Haller D. Dysbiosis in intestinal inflammation: Cause or consequence. Int J Med Microbiol. 2016; https://doi.org/10.1016/j.ijmm.2016.02.010 PMID: 27012594

41. Glasser A-L, Darfeuille-Michaud A. Abnormalities in the handling of intracellular bacteria in Crohn's disease: a link between infectious etiology and host genetic susceptibility. Arch Immunol Ther Exp. 2008; 56: 237–244.

42. Biancheri P, Di Sabatino A, Corazza GR, MacDonald TT. Proteases and the gut barrier. Cell Tissue Res. 2013; 351: 269–280. https://doi.org/10.1007/s00441-012-1390-z PMID: 22427120

43. Pender SLF, MacDonald TT. Matrix metalloproteinases and the gut—new roles for old enzymes. Curr Opin Pharmacol. 2004; 4: 546–550. https://doi.org/10.1016/j.coph.2004.06.005 PMID: 15525541

44. Bamba S, Andoh A, Yasui H, Araki Y, Bamba T, Fujiyama Y. Matrix metalloproteinase-3 secretion from human colonic subepithelial myofibroblasts: role of interleukin-17. J Gastroenterol. 2003; 38: 548–554. https://doi.org/10.1007/s00535-002-1101-8 PMID: 12825130

45. Lawrance IC, Rogler G, Bamias G, Breynaert C, Florholmen J, Pellino G, et al. Cellular and Molecular Mediators of Intestinal Fibrosis. J Crohns Colitis. 2015; https://doi.org/10.1016/j.crohns.2014.09.008 PMID: 25306501

46. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and adaptive immunity in inflammatory bowel disease. Autoimmun Rev. 2014; 13: 3–10. https://doi.org/10.1016/j.autrev.2013.06.004 PMID: 23774107

47. O'Sullivan S, Gilmer JF, Medina C. Matrix metalloproteinases in inflammatory bowel disease: an update. Mediators Inflamm. 2015; 2015: 964131. https://doi.org/10.1155/2015/964131 PMID: 25948887

48. Monteleone G, Caruso R, Fina D, Peluso I, Gioia V, Stolfi C, et al. Control of matrix metalloproteinase production in human intestinal fibroblasts by interleukin 21. Gut. 2006; 55: 1774–1780. https://doi.org/10.1136/gut.2006.093187 PMID: 16682426

49. Helikar T, Kowal B, Rogers JA. A cell simulator platform: the cell collective. Clin Pharmacol Ther. 2013; 93: 393–395. https://doi.org/10.1038/clpt.2013.41 PMID: 23549147

50. Helikar T, Kowal B, McClenathan S, Bruckner M, Rowley T, Madrahimov A, et al. The Cell Collective: toward an open and collaborative approach to systems biology. BMC Syst Biol. 2012; 6: 96. https://doi.org/10.1186/1752-0509-6-96 PMID: 22871178

51. Boldhaus G, Bertschinger N, Rauh J, Olbrich E, Klemm K. Robustness of Boolean dynamics under knockouts. Phys Rev E Stat Nonlin Soft Matter Phys. 2010; 82: 021916. https://doi.org/10.1103/PhysRevE.82.021916 PMID: 20866846

52. Taylor AL, Watson CJE, Bradley JA. Immunosuppressive agents in solid organ transplantation: Mechanisms of action and therapeutic efficacy. Crit Rev Oncol Hematol. 2005; 56: 23–46. https://doi.org/10.1016/j.critrevonc.2005.03.012 PMID: 16039869

53. McManus R. Mechanisms of steroid action and resistance in inflammation and disease. J Endocrinol. 2003; 178: 1–4. PMID: 12844329

54. Coutinho AE, Chapman KE. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. Mol Cell Endocrinol. 2011; 335: 2–13. https://doi.org/10.1016/j.mce.2010.04.005 PMID: 20398732

55. Peters LA, Perrigoue J, Mortha A, Iuga A, Song W-M, Neiman EM, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. Nat Genet. 2017; 49: 1437–1449. https://doi.org/10.1038/ng.3947 PMID: 28892060

56. Laffin M, Madsen KL. Fecal Microbial Transplantation in Inflammatory Bowel Disease: A Movement Too Big to Be Ignored. Clin Pharmacol Ther. 2017; 102: 588–590. https://doi.org/10.1002/cpt.747 PMID: 28695658

57. Fischer S, Neurath MF. Precision Medicine in Inflammatory Bowel Diseases. Clin Pharmacol Ther. 2017; 102: 623–632. https://doi.org/10.1002/cpt.793 PMID: 28699158

58. Dipasquale V, Romano C. Vaccination strategies in pediatric inflammatory bowel disease. Vaccine. 2017; https://doi.org/10.1016/j.vaccine.2017.09.031 PMID: 28967524

59. Danese S, Furfaro F, Vetrano S. Targeting S1P in Inflammatory bowel disease: new avenues for modulating intestinal leukocyte migration. J Crohns Colitis. 2017; https://doi.org/10.1093/ecco-jcc/jjx107 PMID: 28961752

60. Fengming Y, Jianbing W. Biomarkers of inflammatory bowel disease. Dis Markers. 2014; 2014: 710915. https://doi.org/10.1155/2014/710915 PMID: 24963213

61. Helikar T, Konvalina J, Heidel J, Rogers JA. Emergent decision-making in biological signal transduction networks. Proc Natl Acad Sci U S A. 2008; 105: 1913–1918. https://doi.org/10.1073/pnas.0705088105 PMID: 18250321

62. Ben-Horin S, Vande Casteele N, Schreiber S, Lakatos PL. Biosimilars in Inflammatory Bowel Disease: Facts and Fears of Extrapolation. Clin Gastroenterol Hepatol. 2016; https://doi.org/10.1016/j.cgh.2016.05.023 PMID: 27215364

63. van Dullemen HM, van Deventer SJ, Hommes DW, Bijl HA, Jansen J, Tytgat GN, et al. Treatment of Crohn's disease with anti-tumor necrosis factor chimeric monoclonal antibody (cA2). Gastroenterology. 1995; 109: 129–135. PMID: 7797011

64. Present DH, Rutgeerts P, Targan S, Hanauer SB, Mayer L, van Hogezand RA, et al. Infliximab for the treatment of fistulas in patients with Crohn's disease. N Engl J Med. 1999; 340: 1398–1405. https://doi.org/10.1056/NEJM199905063401804 PMID: 10228190

65. Colombel J-F, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Panaccione R, et al. Adalimumab for maintenance of clinical response and remission in patients with Crohn's disease: the CHARM trial. Gastroenterology. 2007; 132: 52–65. https://doi.org/10.1053/j.gastro.2006.11.041 PMID: 17241859

66. Yoshimura N, Yokoyama Y, Matsuoka K, Takahashi H, Iwakiri R, Yamamoto T, et al. An open-label prospective randomized multicenter study of intensive versus weekly granulocyte and monocyte apheresis in active crohn's disease. BMC Gastroenterol. 2015; 15: 163. https://doi.org/10.1186/s12876-015-0390-3 PMID: 26585569

67. Liu Z, Jiang X, Sun C. The efficacy and safety of selective granulocyte and monocyte apheresis for inflammatory bowel disease: A meta-analysis. Eur J Intern Med. 2016; 36: e26–e27. https://doi.org/10.1016/j.ejim.2016.08.028 PMID: 27614377

68. Di Girolamo M, Sartini A, Critelli R, Bertani A, Merighi A, Villa E. Efficacy of a Novel Granulocyte Monocyte Apheresis Adsorber Device in the Treatment of Inflammatory Bowel Diseases: A Pilot Study. Ther Apher Dial. 2016; 20: 668–676. https://doi.org/10.1111/1744-9987.12453 PMID: 27921374

69. Hommes DW, Mikhajlova TL, Stoinov S, Stimac D, Vucelic B, Lonovics J, et al. Fontolizumab, a humanised anti-interferon gamma antibody, demonstrates safety and clinical activity in patients with moderate to severe Crohn's disease. Gut. 2006; 55: 1131–1137. https://doi.org/10.1136/gut.2005.079392 PMID: 16507585

70. Reinisch W, Hommes DW, Van Assche G, Colombel J-F, Gendre J-P, Oldenburg B, et al. A dose escalating, placebo controlled, double blind, single dose and multidose, safety and tolerability study of fontolizumab, a humanised anti-interferon gamma antibody, in patients with moderate to severe Crohn's disease. Gut. 2006; 55: 1138–1144. https://doi.org/10.1136/gut.2005.079434 PMID: 16492717

71. Schreiber S, Fedorak RN, Nielsen OH, Wild G, Williams CN, Nikolaus S, et al. Safety and efficacy of recombinant human interleukin 10 in chronic active Crohn's disease. Crohn's Disease IL-10 Cooperative Study Group. Gastroenterology. 2000; 119: 1461–1472. PMID: 11113067

72. Hueber W, Sands BE, Lewitzky S, Vandemeulebroecke M, Reinisch W, Higgins PDR, et al. Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: unexpected results of a randomised, double-blind placebo-controlled trial. Gut. 2012; 61: 1693–1700. https://doi.org/10.1136/gutjnl-2011-301668 PMID: 22595313

73. Sands BE, Sandborn WJ, Creed TJ, Dayan CM, Dhanda AD, Van Assche GA, et al. Basiliximab does not increase efficacy of corticosteroids in patients with steroid-refractory ulcerative colitis. Gastroenterology. 2012; 143: 356–64.e1. https://doi.org/10.1053/j.gastro.2012.04.043 PMID: 22549092

74. Van Assche G, Sandborn WJ, Feagan BG, Salzberg BA, Silvers D, Monroe PS, et al. Daclizumab, a humanised monoclonal antibody to the interleukin 2 receptor (CD25), for the treatment of moderately to severely active ulcerative colitis: a randomised, double blind, placebo controlled, dose ranging trial. Gut. BMJ Publishing Group Ltd and British Society of Gastroenterology; 2006; 55: 1568–1574. https://doi.org/10.1136/gut.2005.089854 PMID: 16603634

75. Paliogianni F, Ahuja SS, Balow JP, Balow JE, Boumpas DT. Novel mechanism for inhibition of human T cells by glucocorticoids. Glucocorticoids inhibit signal transduction through IL-2 receptor. J Immunol. 1993; 151: 4081–4089. PMID: 8409387

76. Horst HJ, Flad HD. Corticosteroid-interleukin 2 interactions: inhibition of binding of interleukin 2 to interleukin 2 receptors. Clin Exp Immunol. 1987; 68: 156–161. PMID: 3115640

77. El Menyawi M, Fawzy M, Al-Nahas Z, Edris A, Hussein H, Shaker O, et al. Serum tumor necrosis factor alpha (TNF-α) level in patients with Behçet's disease: Relation to clinical manifestations and disease activity. The Egyptian Rheumatologist. 2014; 36: 139–143.

78. Shi M, Wei J, Dong J, Meng W, Ma J, Wang T, et al. Function of interleukin-17 and -35 in the blood of patients with hepatitis B-related liver cirrhosis. Mol Med Rep. 2015; 11: 121–126. https://doi.org/10.3892/mmr.2014.2681 PMID: 25323532

79. Thomas DW, Burns J, Audette J, Carrol A, Dow-Hygelund C, Hay M. Clinical development success rates 2006–2015. San Diego: Biomedtracker/Washington, DC: BIO/Bend: Amplion. 2016;

80. Furfaro F, Gilardi D, Allocca M, Cicerone C, Correale C, Fiorino G, et al. IL-23 Blockade for Crohn s disease: next generation of anti-cytokine therapy. Expert Rev Clin Immunol. 2017; 1–11.

81. Palsson S, Hickling TP, Bradshaw-Pierce EL, Zager M, Jooss K, O'Brien PJ, et al. The development of a fully-integrated immune response model (FIRM) simulator of the immune response through integration of multiple subset models. BMC Syst Biol. 2013; 7: 95. https://doi.org/10.1186/1752-0509-7-95 PMID: 24074340

82. Carbo A, Hontecillas R, Kronsteiner B, Viladomiu M, Pedragosa M, Lu P, et al. Systems modeling of molecular mechanisms controlling cytokine-driven CD4+ T cell differentiation and phenotype plasticity. PLoS Comput Biol. 2013; 9: e1003027. https://doi.org/10.1371/journal.pcbi.1003027 PMID: 23592971

83. Laurence A O'Shea JJ. TH-17 differentiation: of mice and men. Nat Immunol. Nature Publishing Group; 2007; 8: 903–905. https://doi.org/10.1038/ni0907-903 PMID: 17712339

84. Reynolds G, Haniffa M. Human and Mouse Mononuclear Phagocyte Networks: A Tale of Two Species? Front Immunol. 2015; 6: 330. https://doi.org/10.3389/fimmu.2015.00330 PMID: 26124761

85. Mestas J, Hughes CCW. Of mice and not men: differences between mouse and human immunology. J Immunol. 2004; 172: 2731–2738. PMID: 14978070

86. Wasilewska A, Winiarska M, Olszewska M, Rudnicka L. Interleukin-17 inhibitors. A new era in treatment of psoriasis and other skin diseases. Postepy Dermatol Alergol. 2016; 33: 247–252. https://doi.org/10.5114/ada.2016.61599 PMID: 27605893

87. Tan Y, Qi Q, Lu C, Niu X, Bai Y, Jiang C, et al. Cytokine Imbalance as a Common Mechanism in Both Psoriasis and Rheumatoid Arthritis. Mediators Inflamm. 2017; 2017: 2405291. https://doi.org/10.1155/2017/2405291 PMID: 28239238

88. Ryll A, Samaga R, Schaper F, Alexopoulos LG, Klamt S. Large-scale network models of IL-1 and IL-6 signalling and their hepatocellular specification. Mol Biosyst. 2011; 7: 3253–3270. https://doi.org/10.1039/c1mb05261f PMID: 21968890

89. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969; 22: 437–467. PMID: 5803332

90. Kauffman SA. The origins of order: Self organization and selection in evolution. Oxford University Press, USA; 1993.

91. Hopfensitz M, Müssel C, Maucher M, Kestler HA. Attractors in Boolean networks: a tutorial. Comput Stat. Springer-Verlag; 2012; 28: 19–36.

92. Saadatpour A, Albert I, Albert R. Attractor analysis of asynchronous Boolean models of signal transduction networks. J Theor Biol. 2010; 266: 641–656. https://doi.org/10.1016/j.jtbi.2010.07.022 PMID: 20659480

93. Wynn ML, Consul N, Merajver SD, Schnell S. Logic-based models in systems biology: a predictive and parameter-free network analysis method. Integr Biol. 2012; 4: 1323–1337.

94. Saadatpour A, Albert I, Albert R. Attractor analysis of asynchronous Boolean models of signal transduction networks. J Theor Biol. Elsevier; 2010; 266: 641–656. https://doi.org/10.1016/j.jtbi.2010.07.022 PMID: 20659480

95. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R. Modeling systems-level regulation of host immune responses. PLoS Comput Biol. 2007; 3: e109. https://doi.org/10.1371/journal.pcbi.0030109 PMID: 17559300

96. Harvey I, Bossomaier T. Time out of joint: Attractors in asynchronous random boolean networks. Proceedings of the Fourth European Conference on Artificial Life. MIT Press, Cambridge; 1997. pp. 67–75.

97. Willadsen K, Triesch J, Wiles J. Understanding robustness in Random Boolean Networks. ALIFE. 2008. pp. 694–701.

98. Rokach L, Maimon O. Clustering Methods. In: Maimon O, Rokach L, editors. Data Mining and Knowledge Discovery Handbook. Springer US; 2005. pp. 321–352.