THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Pharmacometrics for Combination Therapy in Oncology

Within and Across Species Variability in Time Series and Time-to-event Data

Marcus Baaz

Department of Mathematical Sciences
Division of Applied Mathematics and Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden 2024

Pharmacometrics for Combination Therapy in Oncology
Within and Across Species Variability in Time Series and Time-to-event Data
Marcus Baaz

Department of Mathematical Sciences
Division of Applied Mathematics and Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31 772 1000

Department of Systems and Data Analysis
Fraunhofer-Chalmers Research Centre for Industrial Mathematics
Chalmers Science Park
SE-412 88 Gothenburg
Sweden
Telephone +46 (0)31 309 9400

Typeset with LaTeX
Printed by Chalmers digitaltryck
Gothenburg, Sweden 2024

Pharmacometrics for Combination Therapy in Oncology

Within and Across Species Variability in Time Series and Time-to-event Data

Marcus Baaz

Department of Mathematical Sciences
Division of Applied Mathematics and Mathematical Statistics
Chalmers University of Technology and University of Gothenburg

Abstract

Mathematical modeling is an integral part of the drug development process. Models are developed to describe tumor dynamics or drug concentration to answer questions such as: What concentration is required to reach the desirable treatment outcome? What drug dose and frequency should a drug prescription specify to achieve this concentration? Models are also used for simulation, reducing the need to perform additional animal trials.

In this thesis, we consider how to model dynamical systems that incorporate biologically relevant phenomena such as drug elimination, tumor growth, and the effect of combination therapies consisting of anticancer drugs and radiation treatment. Survival analysis and time-to-event modeling are also discussed as well as how to combine these types of probabilistic models with the dynamical system models in a so-called joint model. All discussed models contain parameters that must be estimated using experimental data and how this estimation is done is considered along with how to deal with variability on different levels, *e.g.*, between individuals and between species.

Furthermore, appended are five papers/manuscripts where this is applied to real-world problems. (I) concerns modeling of radiation therapy in combination with radiosensitizers, (II) presents a translational approach for predicting clinical results using a preclinical model, and (III) focuses on predicting progression-free survival using joint modeling. The last two are in manuscript form and (IV) presents a parametric model for sample size calculations and (V) considers how predictions of progression-free survival are distributed under different models.

Keywords: Mathematical Modeling, Nonlinear Mixed Effects, Pharmacometrics, Oncology, Combination Therapy, Radiation Therapy

List of publications

This thesis is based on the work represented by the following papers and manuscripts:

I. **M. Baaz**, T. Cardilin, F. Lignet, A. Zimmermann, S. El Bawab, J. Gabrielsson, M. Jirstrand, Model-Based Assessment of Combination Therapies – Ranking of Radiosensitizing Agents in Oncology

II. **M. Baaz**, T. Cardilin, F. Lignet, M. Jirstrand, Optimized Scaling of Translational Factors in Oncology - From Xenografts to RECIST

III. **M. Baaz**, T. Cardilin, M. Jirstrand, Model-based Prediction of Progression-free Survival for Combination Therapies in Oncology

IV. **M. Baaz**, T. Cardilin, T. Lundh, M. Jirstrand, Probabilistic Analysis of Tumor Models to Support Early Clinical Trial Design

V. **M. Baaz**, T. Cardilin, M. Jirstrand, Analyzing the Distribution of Progression-free Survival for Combination Therapies: A Study of Model-Based Translational Predictive Methods in Oncology

Authors contribution

I. Implemented the tumor model and the simulation-based TSE method in Mathematica. Performed all computations, further analyzed and refined the model, created all figures and tables, drafted and edited the manuscript.

II. Performed a literature search to find all data used in the project. Implemented the models in Mathematica as well as wrote code for the clinical prediction procedure and algorithm for finding optimized scaling exponents. Performed all computations, created all figures and tables, drafted and edited the manuscript.

III. Performed a literature search to find all data used in the project. Wrote all code for the implementation of the models in Monolix and the progression-free survival prediction algorithm in Mathematica. Performed all computations, created all figures and tables, drafted and edited the manuscript.

IV. Derived the analytical equations. Developed all code for validation in Mathematica. Created all figures and tables, drafted and edited the manuscript.

V. Performed a literature search to find all data used in the project. Derived the analytical results. Implemented all models in Monolix and performed all computations. Created all figures and tables, drafted and edited the manuscript.

Acknowledgements

This PhD project was a collaboration between Fraunhofer-Chalmers Center (FCC), the Department of Mathematical Sciences at Chalmers University of Technology, and Merck. I therefore first want to thank FCC and Chalmers for providing a work environment that has enabled me to carry out the research presented in this thesis. Alongside this, I express my thanks to Merck for supporting the project financially.

Also deserving of my thanks is Torbjörn Lundh for motivating me to apply for this PhD project nearly five years ago. Without your encouragement, I would probably never have pursued this degree. Additionally, I thank you for being my main supervisor from Chalmers and especially for your enthusiasm and excitement, which have always had a positive influence on both me and our research projects.

Equally important, I want to thank Mats Jirstrand who has been my main supervisor from FCC. Thank you for going over my rough manuscript drafts and finding numerous missing parentheses, missing italics, and similar mistakes. I also appreciate all the discussions we have had on more technical matters over the years! I also thank Tim Cardilin who has been the supervisor I have worked closest with during these last years. You have had the pleasure of reading even rougher manuscript drafts than Mats and I have always appreciated your valuable feedback on them.

I would like to acknowledge Johan Gabrielsson for sharing his expertise in the early part of the project. From Merck, I want to thank Floriane Lignet, Samer el Bawab, and Anup Zutshi for their participation in the project. Furthermore, I also thank the Quantitive Pharmacology department for the opportunity to present my research at their forums.

I am thankful as well to all my colleagues at FCC and a special thanks goes out to Viktor and Mattias for making these years go by a bit faster. Finally, I also want to thank my family and friends who unknowingly have contributed to this important research topic!

Marcus Baaz,
Gothenburg, March 2024

Contents

Prologue

This prologue serves as a short introduction to mathematical modeling in pharmacology and aims to give readers who either are clinicians or in similar professions a soft introduction to the mathematics that is discussed in this thesis. Terms such as equations, model parameters, and parameter estimation are explained in a pharmacological context and after reading through this part, the reader has hopefully a slightly better idea of what a mathematician or engineer means with these words.

Ordinary differential equations, or ODEs for short, describe how variables such as drug concentration or distance change with respect to another variable, which often is time, and a mathematical model is often a set of these equations used to portray some real-world scenario. To illustrate, consider a model consisting of an equation detailing how the concentration of an intravenously administered drug changes in patients' blood plasma over time. This equation could take the following form,

$$C(t) = C_0 e^{-\frac{Cl}{V_D}t}, \tag{1}$$

or equivalently,

$$C(t) = C_0 \exp\left(-\frac{Cl}{V_D}t\right). \tag{2}$$

Here C and t are variables that represent the drug concentration in the blood and the time, respectively, and e is Euler's number, which approximately is 2.72. The equation also contains a number of parameters, namely: C_0 the

initial concentration, Cl the clearance, and V_D the volume of distribution. In pharmacological terms, the equation can be seen to represent the purification of the bloodstream from the drug by the liver. Moreover, this equation is the solution to an ODE that we consider in the third chapter of the thesis but for now, it suffices to point out that the equation is known as an exponential function and an illustration of it is shown in Fig. 1

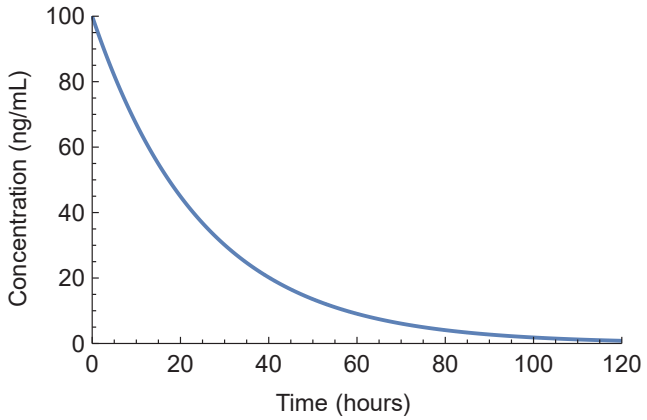

Figure 1: An illustration of an exponential function describing how the drug concentration decreases after an intravenous injection due to the purification of the bloodstream by the liver.

In general, for our model to adequately characterize the concentration dynamics in different organs, processes such as drug absorption, distribution, metabolism, and elimination must all potentially be considered. In the example above, only drug elimination was included, but if the drug instead was given orally, both absorption and distribution would have had to be considered, potentially resulting in the inclusion of more pharmacological parameters and equations. All the parameters included in the model must be estimated using experimental data before the model can be used for the intended purpose. This is known as parameter estimation or model calibration, and in this thesis, we introduce the reader to how these estimations can be performed using varying levels of mathematical sophistication. The need for the most advanced quantitative techniques arises as a result of the high variability often observed in various experiments as well as the wish to minimize sample sizes, *e.g.*, the number of test animals or patients.

Variability can come from many different sources such as different scientists taking measurements, imperfections in the measurement tools, and natural differences between patients. Most mathematical techniques used for analyzing experimental data come from the two closely related fields of probability and statistics and are used to answer questions such as "Is a new treatment

significantly better than the current standard of care?". The word "significant" here, has a special meaning in statistics and the question really posed is, "How certain can we be that the apparent success of one treatment over the other was not caused by random chance". This analysis often relies on *post-hoc* approaches, such as estimating survival curves or performing various statistical tests after the study has been completed. However, the predictive capability of these approaches is limited, *e.g.,* if we modify our question somewhat to be "What is the effect on the survival curve if the treatment schedule changes?" is typically not a question that can be answered using standard statistical techniques. Here is where mathematical modeling provides a powerful predictive approach. By quantifying the relationship between drug concentration, *in-vivo* efficacy, and dropout, a model can be created that can be utilized to *e.g.,* explore how different treatment scenarios are expected to change the trial results. Thus, one purpose of mathematical modeling is to provide a tool for testing hypotheses *in-silico* (in a computer). Hence, it can be an effective approach for reducing the number of real-world experiments that have to be performed. This can both increase the success of treating human patients and also reduce the need for animal testing.

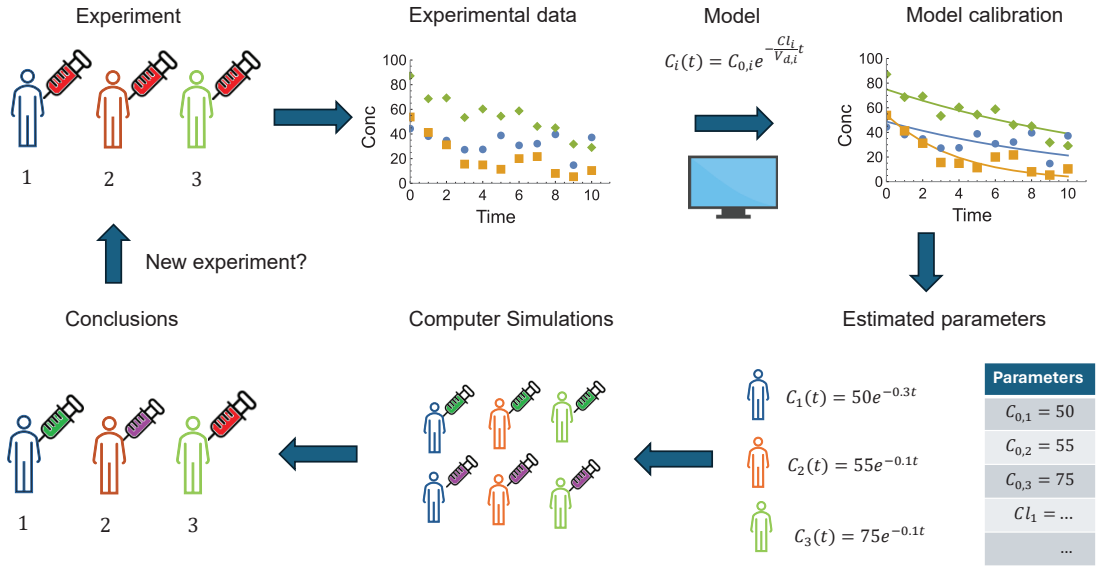

Figure 2: The figure illustrates the modeling process. This example starts with an experiment where a drug is given intravenously to three individuals. Experimental data is obtained by monitoring and repeatedly recording the drug concentration levels at some time point after the injection. A mathematical model is proposed based on prior biological knowledge and by inspecting the data. With the help of a computer, this model is calibrated and parameter estimates are obtained for each individual. By simulating the model, new treatment scenarios, *e.g.*, different dosing schedules can be tested. These simulations can then be used to support conclusions and inform the design of new experiments if such are deemed to be necessary.

1 Introduction

Mathematical modeling is an iterative process that aims to inform researchers about a solution to real-world problems. A research question is posed, such as what is the best dose to give to a patient? Then, a mathematical model, a system of equations, is constructed to allow the phenomena to be studied in a simplified environment [1]. The model can be tested with experimental data to evaluate if it mirrors the real world sufficiently well. If not, the construction and evaluation phases are iterated between until a satisfactory model has been found. The model can then be used to make predictions, which, if possible, should also be validated by new experiments. Through the use of simulations, one can also evaluate if the original question was correctly posed. Perhaps one realizes that what time the drug is given is a crucial aspect that was not originally considered.

Pharmacology, from the Greek words *pharmakon* (drug) and *logia* (knowledge of), is the scientific study of medical drugs and how to medicate patients. Mathematical pharmacology, or pharmacometrics, is the interdisciplinary field where mathematical modeling is used to make inferences in pharmacology [2]. In this thesis, pharmacometrics is used as a guiding tool in the drug development process in oncology, which is the study of tumors or cancer. Cancer is a class of diseases that is characterized by uncontrolled cell growth and metastasis, *i.e.*, spreading to nearby tissues [3], and is a leading cause of death worldwide resulting in approximately one in six deaths [4]. This problem is only predicted to increase as the global population continues to age.

To combat cancer several treatment modalities are standard practice including different types of radiation treatment, surgery, and chemicals. Often a combination of different treatment modalities is required to achieve sufficient high treatment efficacy, *i.e.*, successfully treat a patient. This could take the form of neoadjuvant and adjuvant treatment, where the patient is given an additional treatment before or after the main treatment, respectively. It can also be that the patient is given several treatment modalities simultaneously and this is

called combination therapy. The benefits of this type of treatment can be synergistic effects between the drugs, longer time for the patient to develop drug or radioresistance, and the ability to harness patient variability in response [5, 6, 7]. An example of this is when radiation therapy is combined with a class of anticancer drugs called radiosensitizers. As the name implies, this type of drug causes the tumor tissue to be more sensitive to radiation and thus, a lower radiation dose can be given while still achieving the desired treatment outcome [8, 9]. Another example of a combination therapy currently in use is the concomitant treatment with the two anticancer drugs encorafenib and binimetinib, inhibitors of the BRAF and MEK gene, respectively, for patients with cutaneous melanoma [10].

In the last decades, there has been a growing interest in this field and many anticancer drugs are nowadays used in combination [11]. However, there is still a lack of appropriate tools for identifying drug combinations demonstrating adequate clinical efficacy early in the drug development process [12]. Before anticancer drugs are tested in clinical (human) trials, preclinical (animal or *in vitro*) studies first have to be conducted. Mice implanted with human tumor tissue, commonly referred to as patient-derived xenografts (PDXs), serve as a common choice for this purpose. These mice function as a disease model and, hence, in this context, the word "model" can refer to both a system of equations and a particular PDX. Studies using PDXs are structured in such a way as to closely mirror a clinical trial and two study objectives are to measure how tumor size and drug concentration change over time. These measurements result in time series that often display high variability and appropriate tools are required to analyze these.

A major problem in the drug development process is estimating clinical efficacy from preclinical studies [13, 14]. Commonly, drugs with sufficient preclinical efficacy fail to show similar efficacy in a clinical setting [15]. Although this is the case, studies have also found a correlation between preclinical and clinical efficacy [16]. This highlights the need for new and improved methods of analysis and concepts that can facilitate the translation of preclinical information for clinical use. Mathematical modeling is such a method of analysis that is especially suitable for combination therapies, as all possible combinations cannot be tested experimentally [17].

Compartment models are mathematical models, where each compartment represents a distinct part of a larger system and the movement of some quantity of interest between the compartments is governed by a set of differential equations. This is the go-to type of model in many areas of pharmacometrics and this thesis considers compartment models based on ordinary differential equations (ODEs). These ODEs are formulated using Lipschitz continuous

functions with potential discontinuous jumps at certain times to incorporate treatment schedules. The initial value problems are only considered on fixed time intervals with a single jump occurring at specified times, thus, the existence and uniqueness theorem guarantees a unique solution to them. These solutions are piece-wise continuous functions with potentially a discontinuity at the time of the jumps.

In this thesis, we explore concepts such as pharmacodynamics, time-to-event modeling, and parameter estimation to better understand how they can be used to guide drug development and research. The primary focus of the parameter estimation is on quantifying different forms of variability and how this knowledge can be leveraged to perform more robust model prediction. The focus is on oncology and combination therapies, yet, the methods and techniques have broader applicability beyond this specific context.

1.1 Research Questions

The research discussed in this thesis aims to answer the following questions, in the context of pharmacometrics and oncology.

- How can advanced modeling techniques be used to better optimize study design?
- What is the role/importance of different levels of variability when performing predictions?
- In what way can mathematical modeling provide translational insights?

1.2 Limitations and Scope

The research encompasses preclinical, clinical, and translational aspects with a focus on *in vivo* efficacy. No *in vitro* research is considered and modeling of toxicological effects is only touched upon lightly. Moreover, the dynamical modeling is based on ordinary differential equations and neither partial nor stochastic differential equations are considered.

Furthermore, new and improved algorithms have been developed and implemented, in Mathematica. However, these are not used for the estimation of the model parameters. Instead, previously developed software for this has

been used, including both a Mathematica package developed at Fraunhofer-Chalmers Centre and Monolix [18, 19]. While the thesis delves into the algorithms within these softwares, the level of detail and implementation is constrained.

2 Pharmacometrics

The two essential branches of pharmacology are pharmacokinetics (PK) and pharmacodynamics (PD). In short, PK describes what the body does to the drug, and PD what the drug does to the body [20].

2.1 Pharmacokinetics

PK describes how drugs are absorbed, distributed, metabolized, and eliminated by the body. The aim is often to develop a model that describes how the concentration of a drug changes over time at the target tissue, which is more commonly known as the site of action. Measurements at the site of action are often invasive and it is therefore common to use the concentration in the blood plasma as a proxy. This can be estimated by taking a blood sample and using LC-MS/MS, an analytical technique for detecting compounds in liquid samples by combining liquid chromatography (LC) and mass spectrometry (MS) [21]. Performing such a measurement would give the total drug concentration. However, as a drug enters the bloodstream, a fraction of it binds to the plasma and tissue proteins. These bound drug molecules are not able to interact with the pharmacological target proteins, such as receptors and enzymes, and it is therefore only the unbound drug concentration that is of interest when quantifying the drug's efficacy. Estimating the unbound drug concentration requires additional analysis using methods such as equilibrium dialysis and ultrafiltration [22]. Additional analysis can be expensive and time-consuming, thus, depending on the purpose of the model the total drug concentration (in the blood plasma) might be sufficient to carry out the intended analysis. An example of typical PK time series data is shown in the right part of figure 2.1.

To give a concrete example of a PK compartment model, we consider the case where a patient is prescribed a drug that is to be taken orally and how to

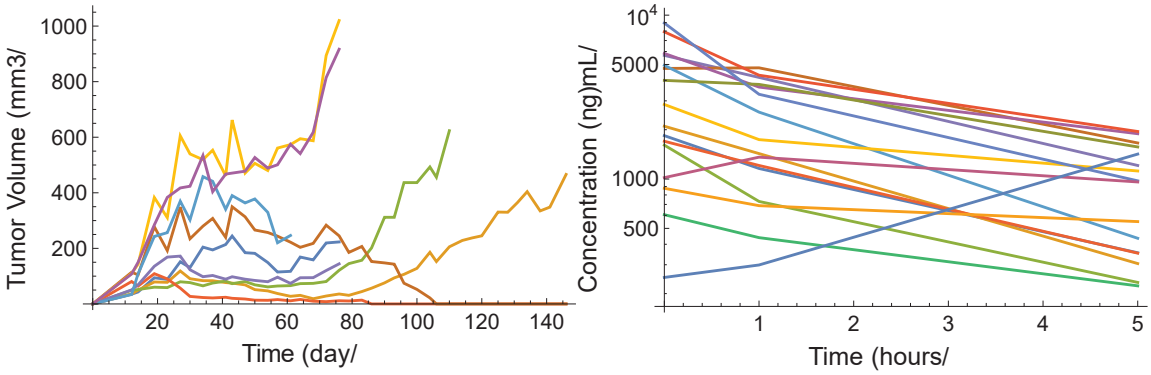

Figure 2.1: Two examples of typical time series data obtained from experiments. Each colored line represents how the tumor volume (left) or drug concentration (right) changes for a specific individual throughout a study.

build a model for the total drug concentration in the blood plasma. Once the drug has been ingested, its first destination is the gut, from whence it is later absorbed into the blood plasma. The blood plasma is constantly being cycled through the liver, where it is purified from toxins and other exogenous (external) substances. Thus, we identify two essential processes that have to be taken into consideration when we attempt to model how the concentration of this drug changes over time in the patient's blood plasma, namely drug absorption and elimination.

In this case, the compartments are the gut and blood plasma and we denote the drug concentration in these compartments as C_A and C_B, respectively. The absorption and elimination can be modeled as first-order processes with rate parameters k_a and k_e, respectively. The system of ODEs that describes this is

$$\begin{aligned} \frac{\mathrm{d}C_A}{\mathrm{d}t} &= -k_a C_A, \quad C_A(0) = \frac{D}{V_d}, \\ \frac{\mathrm{d}C_B}{\mathrm{d}t} &= k_a C_A - k_e C_B, \quad C_B(0) = 0, \end{aligned} \tag{2.1}$$

where D denotes the drug dose and V_d is the volume of distribution and an illustration of the model is shown in Figure 2.2. This system of ODEs can be solved analytically by *e.g.*, first solving the first row separately and then using the technique of integrating factors for the second row.

$$\begin{aligned} C_A(t) &= \frac{D}{V_d} \exp(-k_a\, t), \\ C_B(t) &= \frac{D\, k_a}{V_d(k_e - k_a)} (\exp(-k_a\, t) - \exp(-k_b\, t)). \end{aligned} \tag{2.2}$$

We can see that only a single equation is required to describe the drug concentration in the blood plasma and, therefore, this model is said to be a one-compartment PK model.

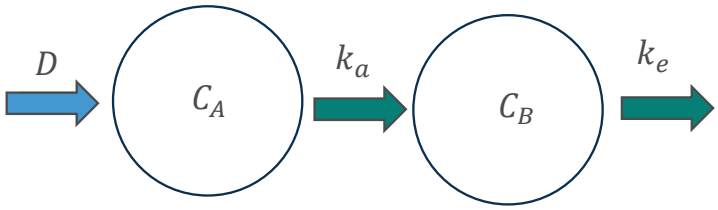

Figure 2.2: An illustration of a one-compartment PK model. D is the drug dose and C_A and C_B are the drug concentration in the gut and blood plasma, respectively. After the drug is ingested, it is absorbed from the gut to the blood plasma from where it is eliminated by the liver. Both processes are modeled as first-order processes with rate parameters k_a and k_e, respectively. This leads to the concentration dynamics being described by the exponential functions shown in Eq. 2.2.

The model is, as all models are, a simplification of what happens in the real world and one important assumption we have made is that the drug is equally distributed through the body. However, this is not necessarily true since some tissues, such as muscle are poorly perfused when compared with *e.g.*, the kidneys [23]. This would lead you to expect the drug concentration to be lower in the muscles in comparison with more perfused tissues. The one-compartment model can easily be extended to account for this by splitting the C_B compartment into two distinct compartments, one describing the concentration in the poorly perfused tissues and another in the well-perfused tissues. We called these two new compartments the peripheral and central compartments and denote the concentration in each by C_P and C_C, respectively. The movement between these two compartments is again modeled as first-order processes with rate parameters k_{12} and k_{21}. The system of ODEs that describes the drug concentration then becomes,

$$\begin{aligned}
\frac{\mathrm{d}C_A}{\mathrm{d}t} &= -k_a C_A, \quad C_A(0) = \frac{D}{V_d}, \\
\frac{\mathrm{d}C_C}{\mathrm{d}t} &= k_a C_A - (k_e + k_{12})C_C + k_{21}C_P, \quad C_C(0) = 0, \\
\frac{\mathrm{d}C_P}{\mathrm{d}t} &= k_{12}C_C - k_{21}C_P, \quad C_P(0) = 0.
\end{aligned} \tag{2.3}$$

In the two models we have investigated, we have assumed that a drug is given with dose D at time $t = 0$. However, to increase the probability of tumor eradication it is more likely that the drug has to be given on multiple occasions. Therefore, another important PK aspect that must be considered is the treatment schedule of the administered drugs. The ODE model presented in Eq. 2.3 can be extended to consider a treatment schedule by including a sum of Dirac delta functions. Consider τ and D to be vectors that represent the time that the drug is given and the corresponding dose. The model is then extended as,

$$\begin{aligned}
\frac{\mathrm{d}C_A}{\mathrm{d}t} &= -k_a C_A + \sum_{n=1}^{N} \frac{D_i}{V_d}\delta[\tau_i - t], \quad C_A(0) = \frac{D_0}{V_d}, \\
\frac{\mathrm{d}C_C}{\mathrm{d}t} &= k_a C_A - (k_e + k_{12})C_C + k_{21}C_P, \quad C_C(0) = 0, \\
\frac{\mathrm{d}C_P}{\mathrm{d}t} &= k_{12}C_C - k_{21}C_P, \quad C_P(0) = 0.
\end{aligned} \tag{2.4}$$

The presented PK models can be used to simulate the concentration profile of drugs, which in turn can be used in the quantification of the drugs' efficacy. When considering drugs given repeatedly, a significant concentration level may still be present in the body when the next dose is given. Therefore, some time may pass before a steady-state level is reached. Simulations of a model can be of use to give an idea of what drug dose to give to achieve a specific steady-state concentration.

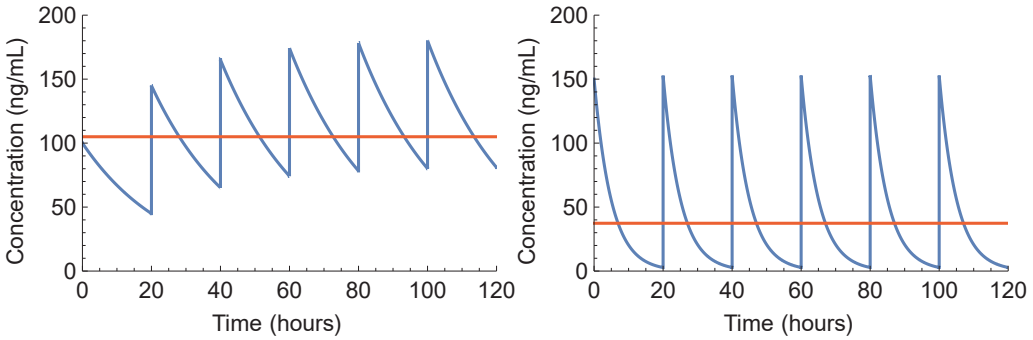

Figure 2.3: Simulations of plasma concentrations for two repeatedly administered drugs. The red line indicates the steady-state average concentration. In the left figure, the drug elimination is not quick enough for the concentration to decrease to zero before the next dose is given. This leads to an initial period where the concentration dynamics differ from the steady-state behavior, the maximum concentration before the second dose is even below the average steady-state concentration. The opposite is true for the illustration in the right figure. Thus, these types of simulations can be useful for determining what dosing schedule to prescribe to achieve a target steady-state concentration.

Depending on the situation it might suffice to use a concentration metric such as the integral of the concentration function *i.e.,* the area under the curve (AUC) or the average concentration instead of the simulated concentration profile. These metrics are easily estimated from the model as,

$$\begin{aligned} AUC_{0-\tau}(\tau) &= \int_0^\tau C_C(x)\,dx, \\ C_{avg}(\tau) &= \frac{AUC_{0-\tau}(\tau))}{\tau}. \end{aligned} \tag{2.5}$$

Again, depending on the elimination rate of the drug, the transient and steady-state metrics might differ and simulations can help find an appropriate dosing schedule to achieve the desired concentration.

2.2 Pharmacodynamics

In oncology, PD typically describes how tumor volumes are affected by a given treatment. Compartment models are often also used here, but the compartments consist of different types of tumor cells, such as proliferating or damaged cells. An example of PD time series data is shown in the left part of figure 2.1.

2.2.1 Tumor Growth Modeling

One of the simplest ways of modeling tumor growth is by assuming that the growth is proportional to the current tumor size, *i.e.*,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = (k_g - k_k)V, \quad V(0) = V_0, \tag{2.6}$$

where k_g and k_k are growth and kill rate parameters, respectively, V is the tumor volume, and V_0 the initial tumor volume [16]. Using the separation of variables method, it can be seen that this results in an exponential growth function with an exact solution,

$$V(t) = V_0 \exp((k_g - k_k)t). \tag{2.7}$$

One flaw of this model is that there is no limitation on how large the tumor can grow and we see that if $k_g - k_k > 0$, $V(t) \to \infty$ as $t \to \infty$, which, of course, is not realistic. Another tumor growth model that solves this is the logistic growth model, where the dynamics are described by,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = (k_g - k_k)(1 - \frac{V}{K_c})V, \quad V(0) = V_0, \tag{2.8}$$

where K_c is the carrying capacity, *i.e.*, the largest sustainable tumor size. Here the tumor initially exhibits exponential growth, but as the size of the tumor increases, the available space and nutrients diminish, resulting in reduced growth. When the tumor volume reaches K_c a steady-state solution is reached.

Another common model with changing growth behavior is the Gompertz growth model, where the tumor cells are assumed to initially be in a phase of exponential growth, which is followed by a phase of linear growth. The model is described by,

$$\begin{aligned} \frac{\mathrm{d}V}{\mathrm{dt}} &= \lambda_0 V, \quad V \leq V_S, \\ \frac{\mathrm{d}V}{\mathrm{dt}} &= \lambda_1, \qquad V > V_S, \\ V(0) &= V_0, \end{aligned} \tag{2.9}$$

where λ_0 and λ_1 are the exponential and linear net growth rate parameters, respectively, and V_S is the tumor size where the switch between the two phases occurs. For computational reasons, it can be useful to use a single ODE to describe the model. The following ODE is useful in approximating the original system,

$$\begin{aligned} \frac{\mathrm{d}V}{\mathrm{dt}} &= \frac{\lambda_0 V}{\left(1 + \left(\frac{\lambda_0}{\lambda_1} V\right)^p\right)^{1/p}}, \\ V(0) &= V_0. \end{aligned} \tag{2.10}$$

These models have all assumed that the tumors are made up of homogenous cells, all behaving in the same manner. However, this is, again, a simplification. Tumors are often seen as having a geometry similar to an ellipsoid with a necrotic core, consisting of dead and dying cells, and outer layers of proliferating cells. Therefore, a more biologically reasonable model differentiates between proliferating and non-proliferating cells. This can be accomplished by introducing a chain of transit compartments to the model [24]. We do this for the exponential growth model, eq. 2.7, and end up with,

$$\begin{aligned}
\frac{dV_1}{dt} &= (k_g - k_k)\, V_1, \\
\frac{\mathrm{d}V_2}{\mathrm{d}t} &= k_k V_1 - k_k V_2, \\
\frac{\mathrm{d}V_3}{\mathrm{d}t} &= k_k V_2 - k_k V_3, \\
\frac{\mathrm{d}V_4}{\mathrm{d}t} &= k_k V_3 - k_k V_4.
\end{aligned} \tag{2.11}$$

$$V_i(0) = V_0 \left(\frac{k_k}{k_g}\right)^{i-1} \qquad i = 1, 2, 3, 4,$$

where V_1 is the volume of proliferating cells, V_2, V_3, and V_4 the volume of non-proliferating cells, k_g the growth rate, and k_k the natural kill rate. The total tumor volume, V_{tot} is given by,

$$V_{tot} = V_1 + V_2 + V_3 + V_4. \tag{2.12}$$

The initial conditions are chosen such that in the absence of treatment, the tumor cells have strictly exponential growth [25]. Furthermore, any number of transit compartments could be considered, but studies have shown that three often is sufficient to explain observed data.

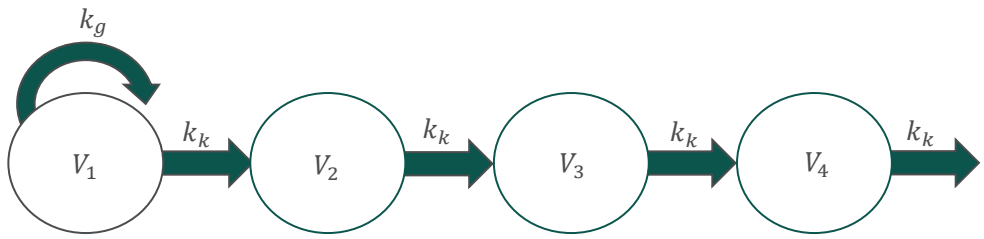

Figure 2.4: An illustration of a compartment tumor model. V_1 contains proliferating tumor cells that are growing and dying with rates k_g and k_k, respectively. Compartments V_2, V_3, and V_4 represent damaged and dying cells that have lost the ability to proliferate. These additional compartments are included based on the biology of tumors and the total tumor volume is the sum of all compartments. Furthermore, they also allow for a delay in treatment effect to be incorporated into the model.

In the next section, we consider how to model treatment effects. For simplicity's sake, we use the exponential growth model as an example, but the same extensions apply to all introduced models. We also note here that the transit compartment models can be used to describe a delay in treatment effect.

2.2.2 Combination of Anticancer Drugs

As previously mentioned, it is only the unbound drug molecules at the target site that can bind to the target protein and have a therapeutic effect. How molecules bind to protein is a well-studied topic in biochemistry and one way of describing the fraction of receptor protein (θ) that is bound by a ligand (L) is,

$$\theta = \frac{L^n}{L^n + K_A^n} = \frac{1}{1 + (\frac{K_A}{L})^n}, \tag{2.13}$$

where n describes the degree of interaction between ligand and binding sites and K_A is the ligand concentration where half of the receptors are bound.

The ligands in this case are the drug molecules, leading to the expectation that there is a nonlinear relationship between drug concentration and efficacy. Thus, when quantifying this relationship the following sigmoid function is useful,

$$E(C_B) = E_{\text{max}} \frac{C_B^n}{EC_{50}^n + C_B^n}, \tag{2.14}$$

where $E_{\max}$ is the maximum efficacy and EC_{50} is the concentration where half of the efficacy is reached. This is generally known as an Emax function and can be seen to be a parameterization of the Hill equation. Another useful sigmoid function is,

$$I(C_B) = 1 - I_{\max} \frac{C_B^n}{IC_{50}^n + C_B^n}, \tag{2.15}$$

where $I_{\max}$ is the maximum inhibition and IC_{50} the concentration leading to a inhibition of 50%. This is known as a maximum inhibitory (Imax) function and is typically used to describe drug inhibition.

It can be important to take the mechanism of action of the modeled drug into consideration when formulating the model. Often a distinction is made between growth inhibitors, such as cetuximab and encorafenib, and cytotoxic agents such as cisplatin and fluorouracil. In general, growth inhibitors function by disrupting the tumor cell's growth by *e.g.,* inhibiting the production of an important protein whereas cytotoxic agents more directly kill the cells by *e.g.,* interfering with DNA replication. For modeling purposes, an Emax function might be a good choice for modeling the efficacy of a cytotoxic drug and an Imax for an inhibitor. A potential model for describing the tumor dynamics when one cytotoxic agent is coadministered with a growth inhibitor is,

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \left(k_g I(C_1) - E(C_2)k_k\right) V, \quad V(0) = V_0, \tag{2.16}$$

where C_1 and C_2 are the blood plasma concentrations of the cytotoxic agent and growth inhibitor, respectively.

As different drugs are given simultaneously there is the possibility of both synergistic (beneficial) and antagonistic (detrimental) combinatory effects. For example, if two concomitant given drugs bind to the same protein receptor they must compete with each other to achieve the binding, potentially resulting in antagonism. These types of interaction effects can be modeled in different ways and one approach is to add an interaction term, that is a function of the concentration of both drugs.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = (k_g I(C_1) - E(C_2)k_k + \gamma(C_1, C_2))\, V, \quad V(0) = V_0. \tag{2.17}$$

Here γ is the interaction function and if we assume it to be a quadratic equation, *e.g.*, $\gamma = \gamma_0 C_1 C_2$, then an antagonistic combination has $\gamma_0 < 0$, a synergistic combination $\gamma_0 > 0$ and an additive combination $\gamma_0 = 0$.

2.2.3 Radiation and Radiosensitizer Treatment

Radiation therapy is another very important treatment modality used against cancers. High doses of ionizing radiation are applied to the area where the tumor is located causing damage to it and surrounding tissue. The most important damage is double-stranded DNA breaks, although single-stranded DNA breaks can also occur. The DNA damage leads to apoptosis and mitotic catastrophe, which are thought to be two of the main responses of tumor cells to irradiation.

The so-called linear-quadratic equation is typically used to quantify cell damage as a result of radiation [26, 27]. The proportion of cells that are damaged after one radiation application with radiation dose, D_R, is given by

$$F(D_R) = 1 - \exp(-\alpha\, D_R - \beta\, D_R^2), \tag{2.18}$$

where α and β are radiosensitivity parameters.

In addition to the more instantaneous effect, repeated exposure to radiation can cause long-term damage to the tumor as well. For example, reductions in growth rates have been observed after longer periods of radiation treatment and it is theorized that this is due to mutations and/or reduced vascularization in the tumor, as well as changes in the tumor microenvironment [28]. One approach to modeling this inhibition after m applications of radiation with dose D_R is once again by using an Imax function,

$$I(D_R) = 1 - I_{\max} \frac{\sum_{i=1}^{m} D_R}{ID_{50} + \sum_{i=1}^{n} D_R}, \tag{2.19}$$

where ID_{50} is the radiation dose leading to a 50% growth inhibititon.

To model both the reduction in growth rate and the direct killing of tumor cells simultaneously the following model could be a choice,

$$\begin{aligned} \frac{\mathrm{d}V}{\mathrm{d}t} &= (k_g I(D_R) - k_k)\, V, t \neq t_i, \quad V(0) = V_0. \\ V(t_i^+) &= V(t_i^-) - F(D_R)\, V(t_i), \quad t = t_i \end{aligned} \tag{2.20}$$

where t_i denotes the time of each radiation application and t_i^- and t_i^+ the time immediately before and after each radiation application, respectively.

Radiosensitizers belong to a class of anticancer drugs that are given in combination with radiation treatment to make tumor tissue more sensitive to radiation [8]. Cells in a hypoxic state are more resistant to radiation treatment and an important class of radiosensitizers, Oxygen Mimetics, function by reducing the population of them [9]. Another important class of radiosensitizers includes those that regulate crucial pathways, such as inhibitors of enzymes involved in the repair mechanism of DNA for both single- and double-stranded breaks [29]. The emergence of nanotechnology has opened up new possibilities in the development of radiosensitizers. Heavy-metal nanomaterials, such as gold and silver particles, show promise as radiosensitizers because of their capacity to absorb, scatter, and emit radiation energy [30, 31]. Moreover, it is possible to tailor the physical characteristics of these particles to, *e.g.*, enhance their accumulation at the tumor target site [32]. Thus, through the utilization of radiosensitizers, it becomes possible to lower the radiation dose to reduce harmful side effects while still achieving a sufficiently high anticancer efficacy.

To model the effect of adding a radiosensitizer to radiation treatment we can extend both the Imax and linear quadratic function in the following manner,

$$\begin{aligned} F(D_R, C) &= 1 - (1 + b\, C) \exp(-\alpha\, D_R - \beta\, D_R^2), \\ I(D_R, C) &= 1 - I_{\max} \frac{\sum_{i=1}^{m} (1 + a\, C) D_R}{ID_{50} + \sum_{i=1}^{m} (1 + a\, C) D_R}, \end{aligned} \tag{2.21}$$

where a and b describe the radiosensitizer's ability to increase the long-term and short-term damages, respectively. For a full dynamical model Eq. 2.20 could again be used after replacing the regular efficacy functions with these extended ones.

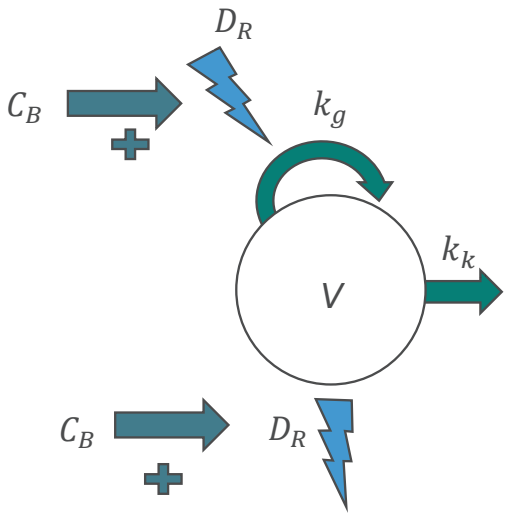

Figure 2.5: A schematic representation of the long-term radiation and radiosensitizer model. V consists of proliferating tumor cells with growth rate and natural kill rate denoted by k_g and k_k, respectively. After each radiation application, with dose D_R, a proportion of proliferating cells are instantly killed. The radiation treatment also inhibits the growth rate. C_B denotes the concentration of radiosensitizer at the instant of radiation application and increases the effect of both types of radiation damage.

2.2.4 Tumor Static Exposure

It is often of interest to find a threshold for drug or radiation exposure that, when surpassed, yields the desired treatment outcomes such as tumor shrinkage. A visual tool for this purpose is the isobologram, which is a curve that describes all exposure pairs with equivalent effects. Tumor Static Exposure (TSE), originally referred to as Tumor Static Concentration is an isobologram that describes all combinations of exposure levels that, when maintained at a constant level, induce tumor stasis, effectively partitioning the space of potential exposures into regions of tumor growth and tumor shrinkage. Administering doses yielding exposures above the TSE curve is predicted to cause tumor shrinkage, possibly leading to complete eradication. The TSE curve for a model is obtained by solving the right-hand side of the tumor volume derivate

equal to zero. The TSE curve based on Eq. 2.16 is shown in Fig 2.6.

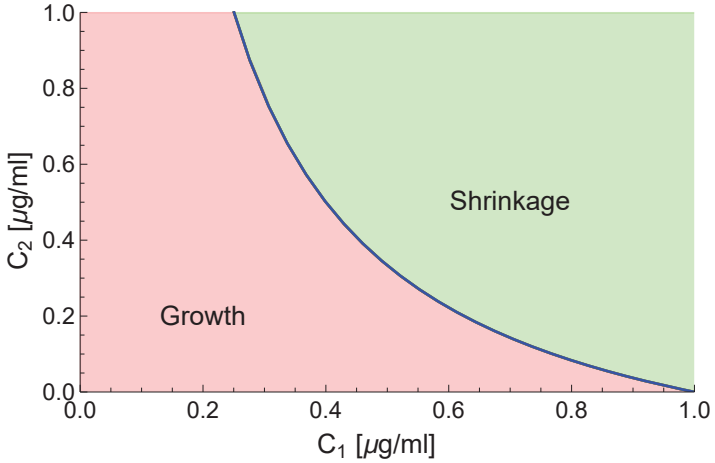

Figure 2.6: An illustration of a TSE curve. Concentration pairs on the blue line are predicted to result in tumor stasis. Hence, concentration pairs below and above the line are predicted to lead to tumor growth and tumor shrinkage, respectively. This can be used as a visual tool for evaluating combination therapies.

An advantage of deriving analytical expressions for the TSE directly from the mathematical model describing the tumor dynamics is that insights are obtained into how various model parameters influence TSE. However, in cases where the tumor dynamics are described by complex models, obtaining analytical expressions for TSE may be impractical without simplifications. A workaround involves turning to numerical methods and simulations.

2.3 Survival Analysis and Time-to-event Modeling

2.3.1 RECIST and Progression-free Survival

Overall survival is defined as the proportion of patients still alive at a certain time point after the start of a clinical trial and is the golden standard for comparing treatments in oncology. However, this metric often takes years to establish and it is therefore important to also have other metrics that can be established earlier.

The Response Evaluation Criteria In Solid Tumors (RECIST) is a framework for assessing tumor size in clinical studies and classifying disease status. The goal of these guidelines is to have a unified method of reporting results from clinical

studies to enable objective comparisons. RECIST states that all tumor lesions have to be accounted for by a combination of target and non-target lesions. The non-target lesions are qualitatively monitored, whereas the target lesions are quantitatively monitored by measuring the sum of the longest diameters of the target lesions (SLD). Based on both of these evaluations the overall response (disease status) of patients can be determined. Overall response is divided into four categories: complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). Treatments could be compared to each other by considering the proportion of patients classified as CR and PR.

Based on SLD measurements CR is achieved if all target lesions disappear, PR if the SLD decreases by 30% from baseline, PD if the SLD increases by 20% and at least 5 mm from the nadir value, and SD otherwise. If non-target lesions are present their status must also be taken into consideration. Furthermore, the appearance of new lesions automatically leads to PD classification.

Besides overall response, another important clinical efficacy metric derived from RECIST is progression-free survival (PFS). PFS describes the proportion of patients not categorized as progressive disease at a specific time point and can be used as a proxy for overall survival.

The models we have previously discussed have all been used to describe longitudinal data. However, to adequately make clinical predictions based on RECIST, our model must be able to handle probabilistic events such as patient death or the appearance of new lesions. In the rest of this section, we consider how this is done using time-to-event modeling.

2.3.2 Time-to-event Modeling

In survival analysis, one is often interested in knowing how many patients have not experienced an event, *e.g.*, disease progression, at different time points. This can be done using time-to-event modeling, where the aim is to quantify the probability that the event occurs as a function of time. If we let T be a random variable and the time of the event, its cumulative distribution function (CDF) is $P_T(t) = p(T < t)$ and describes the probability that the event has occurred before time t. The probability that the event has not occurred must be one minus the CDF, *i.e.*,

$$S(t) = 1 - P(T < t) = p(T > t), \tag{2.22}$$

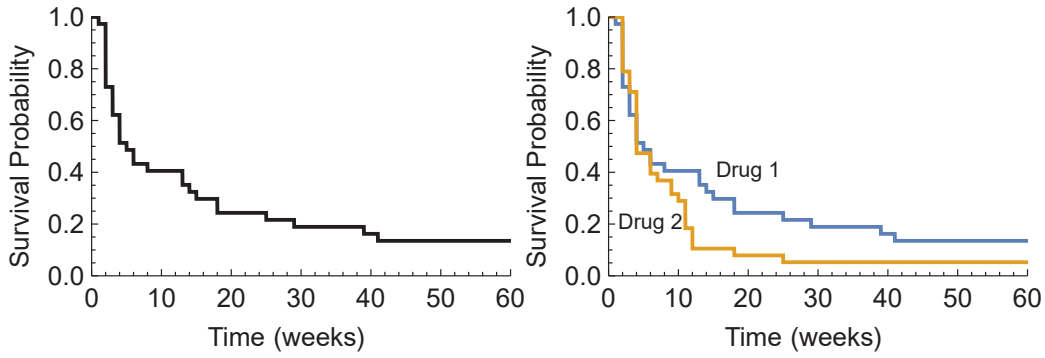

Figure 2.7: (Left) Kaplan-Meier plot showing the probability of survival as a function of time. (Right) Kaplan-Meier plots for two different treatment groups. Drug 1 seems to be more beneficial for patient survival based on this plot. These types of plots can be used to evaluate the efficacy of different drugs.

and this is known as the survival function.

In a real clinical study the actual time of the event is not necessarily known for all patients and to handle this a technique called censoring is used. If a patient leaves the study before the event occurs that patient is said to be right-censored. The idea of right-censoring is that instead of removing patients with unknown event times, they are kept for as long as possible. A nonparametric estimate of the survival function that accounts for this type of censoring is the Kaplan-Meier estimator,

$$\widehat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{2.23}$$

where d_i and n_i are the number of events that occurred and the total population remaining at the i'th time point [33]. The main limitation of this estimator is its ability to handle covariates [34].

Covariates, or patient characteristics, are variables that potentially affect the treatment outcome. A differentiation is made between categorical covariates, such as sex or treatment arm, and continuous covariates such as age or height. To evaluate if women and men respond differently to a drug the Kaplan-Meier estimator could be used to estimate the two groups separately and then compare the resulting survival curves. However, as the number of categorical covariates increases, the sample size of each subpopulation decreases, leading to lower precision of the estimated curves. Furthermore, if continuous covariates, such as age, are of interest, this approach no longer works. This problem is solved by instead using semi-parametric or fully parametric models

to describe the data.

It is sometimes preferable to work with a hazard function, $h(t)$, instead of the survival function because it is generally more informative of the underlying reasons for the occurrence of the event [35]. The hazard function is defined as the event rate at time t conditioned on survival until time t or later, *i.e.*,

$$\begin{aligned} h(t) &= \lim_{\tau \to 0} \frac{\mathrm{p}(t \leq T < t + \tau \mid T > t)}{\tau} \\ &= \lim_{\tau \to 0} \frac{\mathrm{p}(t \leq T < t + \tau)}{\tau} \frac{1}{S(t)} = -\frac{S'(t)}{S(t)}. \end{aligned} \tag{2.24}$$

Any valid hazard function must be non-negative and $\int_0^\infty h(\tau)d\tau = \infty$. The survival function can always be obtained from the hazard function by the following formula,

$$S(t) = \exp\left(-\int_0^t h(\tau)d\tau\right). \tag{2.25}$$

The Cox-proportional hazard model is a semi-parametric approach for survival analysis that takes covariates into consideration. The hazard function is given by,

$$h(t) = h_0(t)\exp\left(X_i\beta\right). \tag{2.26}$$

where X_i are the observed covariates for individual i and β the influence of the covariates on the hazard [36]. h_0 is the baseline hazard and is estimated using a nonparametric estimator, such as the Breslow estimator [37], given by,

$$h_0(t) = \sum_{\forall i : t_i \leq t} \frac{d_i}{\exp\left(X_i\beta\right)}. \tag{2.27}$$

A parametric model could also be used to specify the baseline hazard, *e.g.*, the Weibull model, given by,

$$h_0(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}, \tag{2.28}$$

where k and λ are the shape and scale parameters of the Weibull distribution, respectively.

2.4 Joint Modeling

It may be that the longitudinal data influences the risk for the observed event and vice versa. For example, one might reasonably assume that the probability that a patient dies of cancer increases as the tumor size increases. Such knowledge is something that can be of importance to incorporate into the modeling framework. A sequential approach to this could be to first model the longitudinal data and then *e.g.,* use the estimated tumor doubling time or tumor volume at a certain time point as a covariate in the time-to-event model.

However, one drawback of this approach is that the effect of the event on the longitudinal data is not considered. This can potentially lead to survival bias in the longitudinal model, an overestimation of the precision of the time-to-event model parameters, and a less efficient estimation of both models. To remedy this, a joint modeling approach can be utilized where both models are fit to the data simultaneously. This alleviates the problems with the sequential approach at the cost of more complex computations. The following model is an example of a joint model that potentially could describe the tumor volume dynamics and the probability of disease progression,

$$\begin{aligned} \frac{\mathrm{d}V}{\mathrm{d}t} &= (k_g - k_k)V, \quad V(0) = V_0, \\ h(t) &= \max\left(\alpha + \beta\frac{\mathrm{d}V}{\mathrm{d}t}, 0\right), \end{aligned} \tag{2.29}$$

where $\alpha \geq 0$ is the baseline hazard and $\beta \geq 0$ is the coefficient of the effect of the tumor volume on the hazard. The max function is included to ensure that $h(t) \geq 0$ for all $t \geq 0$.

3 Model Estimation

3.1 Parameter Estimation

All models presented in the previous chapter contain unknown parameters that must be estimated using experimental data. To do so requires an estimator, which is a function of the data whose output are estimates of the model parameters, θ^*. The most commonly used estimator in this setting is the maximum likelihood estimator (MLE). This method involves constructing a likelihood function, representing the probability of the data given the observed data, as a function of the model parameters. Subsequently, the goal is to identify the parameters that maximize the likelihood.

3.1.1 Population Modeling

An experiment typically involves more than a single individual and we consider three approaches with increasing sophistication for tackling parameter estimation when considering the population aspect. The first method is the naïve pooled approach where it is assumed that all data comes from a single "hyper" individual that represents the population in large.

We consider a population of N subjects and denote all observed data for individual $i = 1, .., N$ at each discrete time point t_j, $j = 1, ..., M$ by the vector $y_i = (y_{i,1}, y_{i,2}, ..., y_{i,M})$. Furthermore, we let the matrix $Y = (y_1, y_2, ..., y_N)$ contain all observations for all individuals. The aim is to describe the data using a deterministic output term (model prediction) and a residual error,

$$y_i = g(x_i, t, \theta, u_i) + \sigma\, e_i, \tag{3.1}$$

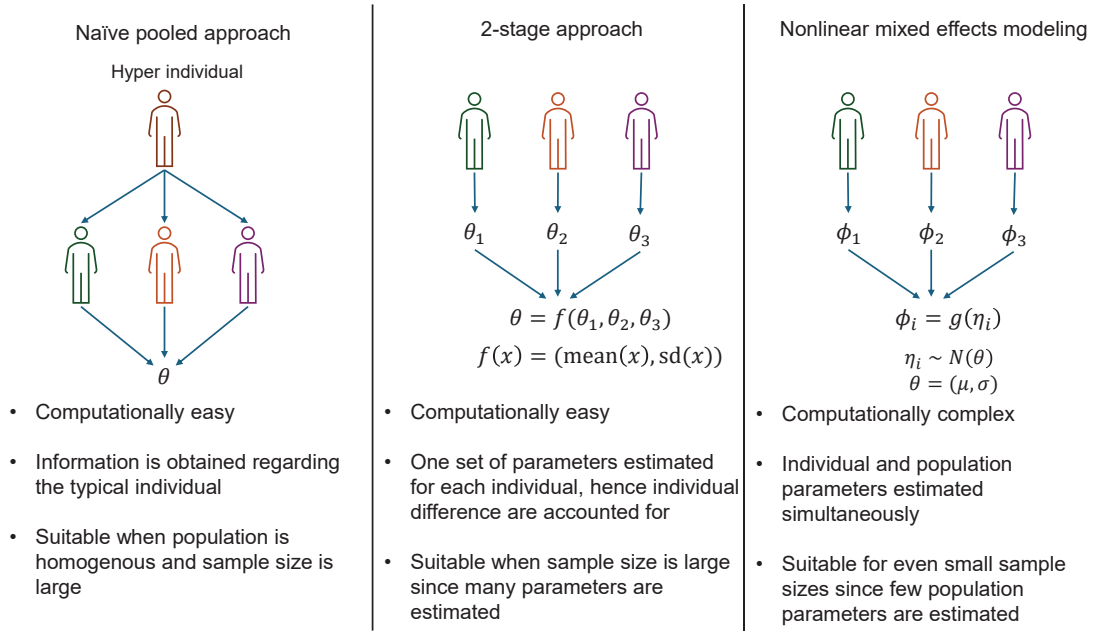

Figure 3.1: Illustration of the naïve pooled approach, 2-stage approach, and nonlinear mixed effects (NLME) modeling for parameter estimation. The naïve pooled approach assumes that all patients are replicates of a "hyper" individual that represents the population at large. This allows the general behavior of the population to be investigated, but limited information regarding the variability in the population is obtained. In the 2-stage approach, a model is first fit to each individual and inferences regarding the population variability are then obtained by analyzing the individual models' parameters. In the NLME framework, population variability is directly incorporated into the model equations. Both random effects (individual parameters) and fixed effects (population parameters) are estimated simultaneously, which allows smaller sample sizes to be considered in comparison with the other approaches. This comes at the cost of both computational and theoretical complexity.

where e_i is a standard normal random vector, *i.e.*, $e_{i,j} \sim N(0,1)$ and σ the standard deviation of the residual errors. g could be one of the previously explored tumor models and θ are the parameters of the model. x are state variables such as the tumor volume itself or the drug concentration and u_i are patient covariates.

Because of the assumptions of the error model, each observation is also normally distributed, *i.e.*, $y_{i,j} \sim N(g(x_{i,j}, t_j, \theta, u_{i,j}), \sigma)$. From the probability density function (PDF) of the normal distribution, we have that the likelihood function is,

$$L(\theta) = \prod_{i=1}^{N} p(y_i|\theta) = \prod_{i=1}^{N}\prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{i,j} - g(x_{i,j}, t_j, \theta, u_{i,j}))^2}{2\sigma^2}\right). \tag{3.2}$$

In this model formulation, the covariates are the only thing that differs between patients, but this is usually not enough to explain the variability stemming from complex biological processes. Thus, all the (unexplained) inter-individual variability (IIV) is confounded with the measurement error. Hence, the knowledge gained regarding the population as a whole is limited. This is the main drawback of the naïve pooled approach. If there are many samples from a homogenous population and the median behavior of the population is of interest it can still be useful as it is an easy and computationally inexpensive approach.

The next method we consider, the two-staged approach, solves the problem of confounding IIV and measurement errors by estimating a set of model parameters for each individual separately. Each observation is now described by,

$$y_i = g(x_i, t, \theta_i, u_i) + \sigma\, e_i, \tag{3.3}$$

where θ_i are the parameters specific to individual i. The first stage consists of maximizing the total population likelihood,

$$L(\theta) = \prod_{i=1}^{N} p(y_i|\theta_i) = \prod_{i=1}^{N}\prod_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_{i,j} - g(x_{i,j}, t_j, \theta_i, u_{i,j}))^2}{2\sigma^2}\right). \tag{3.4}$$

Once accomplished, the second stage is to estimate, *e.g.,* the expectation and variance of θ and use these estimates to describe the IIV. This approach is suitable for a heterogeneous population but a large sample size is required to get accurate estimates of the expectation and variance of θ.

The third and final approach, nonlinear mixed effects (NLME) modeling, alleviates the need for a large sample size by directly estimating the population distribution of the parameters from the data. This is done by considering two types of model parameters, fixed effects parameters, θ, that are the same for the entire population, and random effects parameters, ϕ, that are specific to each individual but are assumed to be drawn from the same distribution.

In the NLME framework, we consider predictive models based on ODEs (or Stochastic ODEs) that can be written in the following way,

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f(x_i, t, u_i, \theta, \phi_i), \quad x_i(0) = r_i(\theta, \phi_i). \tag{3.5}$$

Let $g_i(x_{i,j}, t_{i,j}, u_{i,j}, \theta, \phi_i)$ be the solution to Eq.3.5

To make the following steps a bit more concrete we consider the exponential tumor growth model, eq.2.7, as an example. The predictive model has three parameters, V_0, k_g, and k_k, and if we assume that all three follow lognormal distributions then $\phi_i = (V_{0,i}, k_{g,i}, k_{k,i}) \sim LN(\mu, \Omega)$ and $\theta = (\mu, \Omega, \sigma)$. Here μ is the mean vector and Ω is the covariance matrix of the multi-lognormal distribution. Thus, under the assumption of additive normal error, the observations for individual i are described by,

$$y_i = V_{0,i} \exp\left((k_{g,i} - k_{k,i})t\right) + \sigma\, e_i, \tag{3.6}$$

Recall, if $X \sim N(\mu, \sigma)$, then $Y = \exp(X) \sim LN(\mu, \sigma)$. Thus, we can equivalently consider the model,

$$y_i = \exp(\eta_i(1)) \exp\left(\exp(\eta_i(2)) - \exp(\eta_i(3))\right)t) + \sigma\, e_i, \tag{3.7}$$

where η_i denotes the random effect parameters and follow a multivariate normal distribution.

The problem in this model formulation is that the random effects parameters are not an observed quantity, *i.e.*, they are latent variables and, thus, the likelihood function cannot be evaluated given only θ. A solution to this is to marginalize the individual likelihoods over ϕ_i (or η_i), *i.e.*,

$$L(\theta) = \prod_i p(y_i \mid \theta) = \prod_i \int p(y_i, \phi_i \mid \theta) d\phi_i. \tag{3.8}$$

Using the definition of the conditional probability, we have that,

$$\begin{aligned} L(\theta) = \prod_i \int p(y_i, \phi_i \mid \theta) \mathrm{d}\phi_i = \prod_i \int \frac{p(y_i, \phi_i, \theta)}{p(\theta)} \mathrm{d}\phi_i = \\ \prod_i \int p(y_i \mid \phi_i, \theta) \frac{p(\phi_i, \theta)}{p(\theta)} \mathrm{d}\phi_i = \prod_i \int p(y_i \mid \phi_i, \theta) p(\phi_i \mid \theta) \mathrm{d}\phi_i, \end{aligned} \tag{3.9}$$

where,

$$p(y_i \mid \phi_i, \theta) = \prod_j \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(y_{i,j} - V_{0,i} \exp(k_{g,i} - k_k)t)^2}{2\sigma^2}), \tag{3.10}$$

and the PDF of a multivariate lognormal random vector is,

$$p(\phi_i \mid \theta) = \frac{\prod_k \phi_i(k)^{-1} \exp\left(-\frac{1}{2}(\log(\phi_i) - \mu)^T \Omega^{-1} (\log(\phi_i) - \mu)\right)}{(2\pi)^{n/2} \det(\Omega)^{1/2}}. \tag{3.11}$$

The integral in Eq. 3.9 has to be solved for each individual, but considering Eqs.

3.10 and 3.11, we quickly realize that this integral lacks a closed-form solution and is intractable for even this relatively simple model. In the two upcoming sections, we consider two algorithms for solving this and obtaining estimates of the parameters.

Once the most likely fixed effects parameters have been estimated, θ^*, the most likely individual parameters, *i.e.*, the mode of the conditional random effects density can be found by solving,

$$\phi_i^* = \underset{\phi_i}{\operatorname{argmax}}\, p(\phi_i \mid y_i, \theta^*) = \underset{\phi_i}{\operatorname{argmax}}\, p(y_i \mid \phi_i, \theta^*) p(\phi_i \mid, \theta^*). \tag{3.12}$$

These values are commonly known as Empirical Bayes Estimates (EBEs) and in the η parametrization, the equivalent values are $\eta_i^* = \log(\phi_i^*)$.

3.1.2 First Order Conditional Estimation

The Laplace Approximation, approximates the integral in Eq. 3.9 using a second-order Taylor approximation of the individual log-likelihoods, $l_i = \log(L_i)$. We consider the η parametrized model and perform the Taylor expansion in η around η^* *i.e.*,

$$l_i(\eta_i) \approx l_i(\eta_i^*) + \nabla l_i(\eta_i^*)(\eta_i - \eta_i^*) + \Delta l_i(\eta_i^*) \frac{(\eta_i - \eta_i^*)^2}{2} \tag{3.13}$$

Since $\nabla l_i(\eta_i^*) = 0$, we have,

$$L(\theta) \approx L_L(\theta) = \Pi_i \exp(l_i(\eta_i^*)) \int \exp\left(\frac{(\eta_i - \eta_i^*)^2}{2\Delta l_i(\eta_i^*)^{-1}}\right) d\eta_i, \tag{3.14}$$

where L_L is the approximate total population likelihood. We note that inside the integral is the unnormalized PDF of a multivariate normal random vector with mean $\mu = \eta_i^*$ and covariance matrix $\Sigma = -\Delta l_i(\eta_i^*)^{-1}$, which evaluates to the inverse of the normalization constant of the PDF, *i.e.*, $-\det\left[\frac{-\Delta l_i(\eta_i^*)}{2\pi\eta_i^*}\right]^{-1/2}$. Hence, we obtain the following approximation of the individual log-likelihoods,

$$L_i(\theta) \approx \exp(l_i(\eta_i^*)) \det \left[\frac{-\Delta l_i(\eta_i^*)}{2\pi} \right]^{-1/2}. \tag{3.15}$$

The idea for finding the MLE of θ is by iteratively first finding the η_i^* for each individual that maximizes 3.15, given θ, and then given η_i^* finding the θ that maximizes 3.15. For the first iteration, a start guess of θ needs to be specified. The algorithm is presented below.

Algorithm 1 Maximum Likelihood Estimates of θ

Set initial values $\theta^{(1)}$ and $k = 1$
repeat
  Set $\eta_i^* = \text{argmax}_\eta \, l_i(\eta, \theta^k)$ for each individual
  Set $\theta^{k+1} = \text{argmax}_\theta \sum l_i(\eta_i^*, \theta)$
  $k = k + 1$
until convergence of θ

To compute 3.15 we have to differentiate l_i twice with respect to η_i, which can be a computationally expensive task. The first-order conditional estimation with interaction (FOCEI) approximates the Hessian matrix by ignoring second-order derivatives whereas the first-order conditional estimation (FOCE) method also ignores the dependence of the residual covariance matrix on the random effect parameters. The rationale behind ignoring the second-order terms is that their expected value in a correctly specified model is 0. For both a detailed derivation of the approximations of the Hessian matrix as well as a more detailed MLE algorithm using FOCE or FOCEI we refer the reader to Almquist (2014) and Leander (2021) [18, 38]. Finally, we also mention that there are no guarantees that this algorithm converges, but it often does.

3.1.3 Stochastic Approximation Expectation Maximization

The second algorithm for NLME modeling that we consider is the Stochastic Approximation Expectation Maximization (SAEM) algorithm. SAEM is an extension to the Expectation Maximization (EM) algorithm and it is therefore appropriate to first give a quick overview of this algorithm to understand why it cannot be applied to the problem at hand. The EM algorithm can be used to maximize a similar likelihood as in 3.8, but with a discrete latent variable, z_i. We consider how to maximize the following likelihood,

$$L(\theta) = \prod_i p(y_i \mid \theta) = \prod_i \int p(y_i, z_i \mid \theta) dz_i. \tag{3.16}$$

As previously seen this integral can be difficult to solve and the idea of the EM algorithm is that instead of maximizing this likelihood we iteratively solve,

$$\begin{aligned} Q(\theta|\theta^k) =& E_{z_i \sim p(.|y_i, \theta^{(k)})} \sum_i \log(p(y_i, z_i \mid \theta)) \\ =& \sum_k \sum_i \log p(z_i = k \mid y_i, \theta^{(k)}))(p(y_i, z_i = k)k \mid \theta)), \end{aligned} \tag{3.17}$$

and

$$\theta^{(k+1)} = \underset{\theta}{\operatorname{argmax}}(Q(\theta|\theta^k)). \tag{3.18}$$

The rationale behind this is that it can be shown that,

$$\log(p(y_i) \mid \theta) - \log(p(y_i) \mid \theta^{(k)}) \geq Q(\theta|\theta^k) - Q(\theta^{(k)}|\theta^k), \tag{3.19}$$

i.e., the likelihood function increases with at least as much as $Q(\theta \mid \theta^k)$ for each iteration and thus maximization of Q implies maximization of the likelihood.

Now, the reason why we cannot apply the EM algorithm straight away to maximize 3.8 is that η_i is a continuous random vector and thus we end up with new, often intractable, integrals,

$$Q(\theta|\theta^k) = \sum_i \int p(\eta_i \mid y_i, \theta^k) \log(p(y_i, \eta_i \mid \theta)) d\eta_i. \tag{3.20}$$

However, instead of exactly solving the integrals we can use Monte Carlo integration to make stochastic approximations of them, that is $Q(\theta|\theta^k) \approx Q_L(\theta|\theta^k)$,

with,

$$Q_L(\theta|\theta^k) = \frac{1}{L}\sum_{l=1}^{L}\log(p(y_i, \eta^*_{i,l} \mid \theta) = \frac{1}{L}\sum_{l=1}^{L}\log(p(y_i, \mid \theta, \eta^*_{i,l})\log(\eta^*_{i,l} \mid \theta), \tag{3.21}$$

where $\eta^*_{i,l}$ are samples drawn from the conditional distribution, $p(\eta_i \mid y_i, \theta^k)$. This conditional distribution often does not have a closed form but samples from it can be generated using Markov Chain Monte Carlo (MCMC) methods, *e.g.,* the Metropolis-Hastings algorithm. Here L is the total number of Markov chains drawn for each individual and after drawing these samples eq. 3.21 can be maximized in terms of θ. To achieve a sufficiently stable stochastic approximation a rule of thumb is that the total number of MCMC samples ($N \times L$) should be at least 50.

To give the SAEM algorithm a chance to explore the parameter space well, it starts in what is known as the exploratory phase. Here the estimate of θ at iteration $k = 1, ..., K-1$ is entirely based on the current MCMC sample, *i.e.,*

$$\theta^k = \underset{\theta}{\operatorname{argmax}}\, Q_L(\theta|\theta^{k-1}). \tag{3.22}$$

This means that the algorithm does not retain any information from the previous iterations. However, to ensure convergence, this changes as it enters the smoothing phase at $k = K-1$. Here the estimates are found by averaging over previous $k \geq K$ iterations θ-estimates and the estimate from the current MCMC sample,

$$\theta^k = \theta^{k-1} + \frac{1}{k}(\underset{\theta}{\operatorname{argmax}}\, Q_L(\theta|\theta^{k-1}) - \theta^{k-1}). \tag{3.23}$$

Both phases can be described by,

$$\theta^k = \theta^{k-1} + \frac{1}{k^\alpha}(\underset{\theta}{\operatorname{argmax}}\, Q_L(\theta|\theta^{k-1}) - \theta^{k-1}), \tag{3.24}$$

where α is known as the stepsize exponent and determines the memory of the algorithm. Setting it to 0 and 1 gives eq. 3.22 and eq. 3.23, respectively.

The steps in the SAEM algorithm are summarized below and for a detailed proof of the convergence of the SAEM algorithm, we refer the reader to Delyon *et al.* [39].

Algorithm 2 Maximum Likelihood Estimates of θ using SAEM

Set initial values $\theta^{(1)}$ and $k = 1$
repeat
  for i=1:N **do**
    Generate L samples from $p(\eta_i \mid y_i, \theta^{k-1})$ using MCMC
  end for
  if $k < K - 1$ **then**
    $\alpha = 0$
  else
    $\alpha = 1$
  end if
  $\theta^k = \theta^{k-1} + \frac{1}{k^\alpha}(\text{argmax}_\theta \, Q_L(\theta|\theta^{k-1}) - \theta^{k-1})$,
  $k = k + 1$
until convergence of θ

Since SAEM uses samples from the conditional parameter distributions for estimation, it necessitates that such distributions exist, meaning that only fixed effect parameters that are associated with one of the random effect parameters or the measurement model can be estimated. A workaround for this is to either introduce small artificial variability to the parameter or run a separate maximization algorithm in parallel with SAEM, such as the Nelder-Mead.

3.2 Model Selection and Validation

3.2.1 Parameter Uncertainty

One of the first important tasks when validating a model is to ascertain how certain we can be in the correctness of the estimates, *i.e.*, what is the variance of the estimates? The observed Fisher information matrix is given by,

$$I(\hat{\theta}) = -\frac{d^2 \log(p(Y \mid \theta)}{d\theta^2}\bigg|_{\theta=\theta^*}, \tag{3.25}$$

and describes the curvature of the log-likelihood function in the mode of the distribution. Thus, it gives a qualitative picture of how well the parameters have been estimated. A large absolute value indicates that the mode of the distribution is located in a high-density peak and that a small change of the parameter results in a large reduction of the likelihood. This in turn means that the parameter is most likely well estimated. On the other hand, a small absolute value indicates that the mode is located in a flat-density area and, thus, most likely less well estimated. This is illustrated in Figure 3.2.

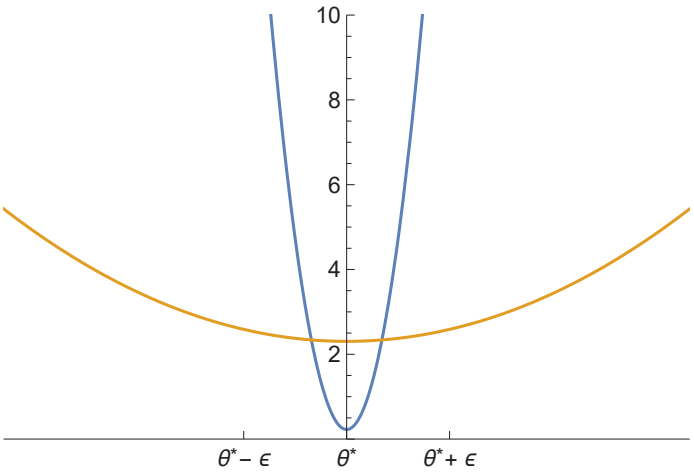

Figure 3.2: The plot shows two different negative log-likelihood functions that both have their minimum value in θ^*. The blue log-likelihood has a very high curvature meaning that a small step ϵ in θ causes a large increase in its value, with the opposite being true for the yellow log-likelihood. From this, we can infer that θ^* in the model with the blue log-likelihood is better estimated in comparison to the θ^* in the model with the yellow log-likelihood.

Moreover, the Cramér-Rao lower bound (CRLB) is the lowest variance an unbiased estimator can obtain and is defined as the reciprocal of $I(\theta^*)$. Maximum likelihood estimators are efficient estimators, *i.e.*, they achieve the CRLB when the sample size tends to infinity, hence we have equality in the equation above. The variance of the estimate of k'th parameter can thus be approximated with $I(\theta_{kk})^{-1}$ when the sample size is sufficiently large. The precision of the parameter estimates is usually presented as relative standard errors % (RSE) and are estimated as,

$$RSE_k\% = 100\% \frac{I(\theta^*_{kk})^{-1/2}}{\theta^*} \tag{3.26}$$

A rule of thumb is that an estimated parameter with an RSE below 25% is well estimated and up to 50% is acceptable.

3.2.2 Validation and Selection

Before a calibrated model can be used for predictions, its ability to adequately describe the observed data must be validated. More specifically, we have made the following three assumptions during the building of the model that must be evaluated: (1) The deterministic function h can approximately describe the data. (2) The residuals, $\hat{e}_{i,j} = y_{i,j} - h(t_{i,j}, ...)$, approximately follows the chosen error model. (3) The individual parameters (EBEs) can be described by the chosen distributions.

These assumptions are often assessed by visually inspecting different plots, but the empirical distributions could also be quantitatively evaluated using *e.g.,* the Kolmogorov–Smirnov test. Plotting the observed data versus the model predictions for the entire population as well as for each individual gives a hint of how well the model fits on a population and an individual level, respectively. Fig. 3.3 shows examples of these kinds of plots.

A plot showing the estimated residuals against both time and the state spaces (*e.g.,* the tumor size) allows one to evaluate if any dependencies have not been accounted for in the model specification. We show examples of such plots from two models in Fig. 3.4. In the left plot, the residual does not seem to be dependent on time, whereas in the middle plot, the residuals seem to increase with time, indicating some misspecification. The residuals could also be plotted together with the PDF of the estimated distributions and under the assumption of normally distributed residuals, a QQ-plot could also be appropriate. For the EBEs, it is important to verify that there are no obvious differences in distribution between treatment groups, which could, *e.g.,* be a sign of drug interaction effects. To test this the η EBEs could, *e.g.,* be plotted separately for each treatment group to evaluate if they all are approximately symmetric around the mean. This is illustrated in the right part of Fig. 3.4, where treatment arm three does not behave like the other arms, which again indicates a misspecification of the model. Histograms with theoretical PDFs and QQ plots can be used to evaluate the entire set of EBEs as well.

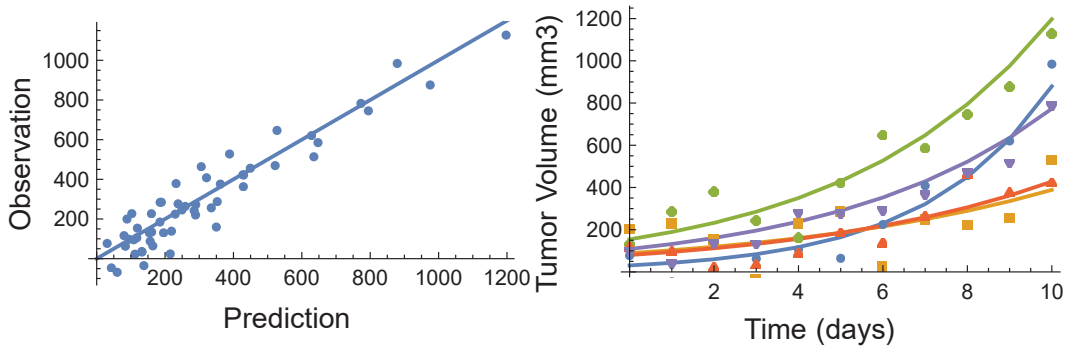

Figure 3.3: (Left) Tumor volume observations plotted against model predictions. Dots on the blue line indicate a perfect prediction. (Right) Individual model predictions (lines) plotted together with the individual data (dots).

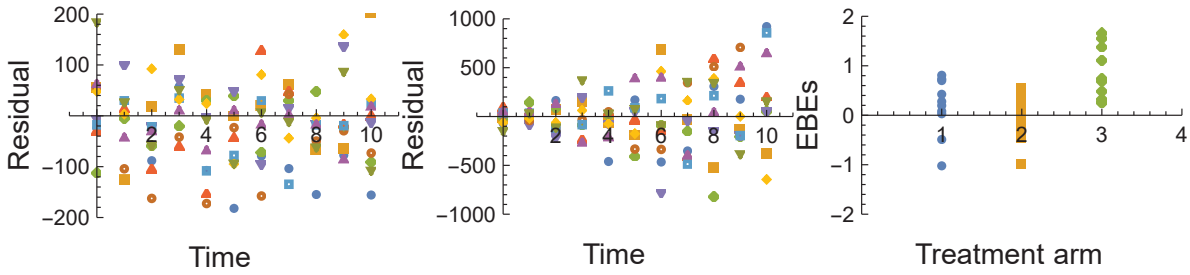

Figure 3.4: (Left) Residuals plotted against time for a correctly specified model. Here the variance seems to be constant (homoscedasticity) and the residuals are approximately symmetrical around 0. (Middle) Residuals plotted against time for a misspecified model. Here the variance of the residuals increases with time (heteroscedasticity), which indicates that a different error model might be more suitable. (Right) EBEs plotted for three treatment arms. Here the third arm seems to differ from arms one and two, potentially caused by a misspecified interaction term.

A popular method for evaluating the overall fit of the model is by performing visual predictive checks (VPCs) [40]. Using this technique, new datasets are generated by simulating the calibrated model. In a regular VPC, the dynamics of the *e.g.*, 10th, 50th, and 90th percentile virtual individuals are estimated and since several datasets are generated, confidence intervals for these percentiles can also be obtained. The simulated percentiles can then be compared with the same percentiles from the experimental data.

Kaplan-Meier VPCs, are similar to regular VPCs but are used for evaluating time-to-event models. Here, datasets are again simulated using the model, and a survival curve is estimated using the Kaplan-Meier estimator for each dataset. Survival percentiles can then be chosen for time points of interest and linear

interpolation use to construct the plot. Examples of both types of VPCs are shown in Fig 3.5.

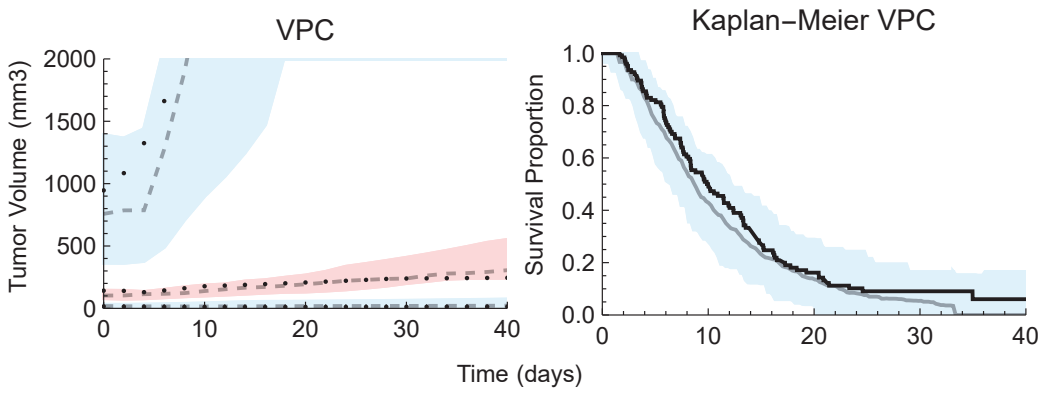

Figure 3.5: (Left) VPC for a dynamical model. The black dots are observed data and grey dashed lines are the predicted 10th, 50th, and 90th percentile virtual individuals. Colored areas are 95% confidence intervals for the predictions. (Right) Kaplan-Meier VPC. The black line is the observed data, the grey is the median model prediction, and the blue area is a 95% confidence interval of the prediction. Both of these types of plots give a good indication of the overall fit of the model.

Several models could look equally good after using this toolbox of validation methods. If the models contain the same number of parameters the one with the lowest likelihood value should be chosen. If the number of parameters differs one often uses Aikake Information Criterion (AIC) or Bayesian Information Criterion which considers the likelihood but also penalizes models with many parameters. They are given by,

$$\begin{aligned} AIC &= 2k - 2\log(L(\theta^*)), \\ BIC &= k\log(n) - 2\log(L(\theta^*)), \end{aligned} \tag{3.27}$$

where k is the number of estimated parameters and n is the number of observations.

3.3 Additional Levels of Variability

3.3.1 Inter-Study Variability

As studies are carried out by different researchers, at different times, or with different study designs, it is possible to get considerable differences in data even for the same experimental setup [41, 42]. This is known as inter-study variability (ISV) or inter-occasional variability (IOV) and the quantification of it can be essential for the assumptions in the model to be valid.

To apply ISV to the k'th fixed effects parameter one estimates a unique value of the parameter for each j'th study (or occasion), *i.e.*,

$$\theta_{k,i} = I_{z_i=j}\theta_k^j \tag{3.28}$$

where $I_{z_i=j}$ is an indicator function that is 1 if individual i is in study j and 0 otherwise. $\theta_{k,i}$ can hence be seen as a multinomially distributed random variable. Moreover, if this parameter is associated with the distribution of the m'th random effect, we can equivalently see that random effect as following a mixture distribution with density given by,

$$p(\phi_{m,i}) = \Sigma_j p(z_i = j)p(\phi_{m,i}^j), \tag{3.29}$$

with $\phi_{m,i}^j \sim f(\theta_k^j, ...)$.

3.3.2 Inter-species Variability

Another type of variability, besides the types discussed in the previous chapter, that is very important when developing drugs, is inter-species variability. Differences in shape, anatomy, and physiology lead species to react differently to the same drug [43]. In oncology, this can, *e.g.*, be differences in drug exposure or tumor growth rate. Translating information from preclinical studies for clinical use is a tough challenge and it is thought that insufficient knowledge in this field is a major contributing factor to the high attrition rates seen for anticancer drugs [14]. Currently, translation from animals to humans is mainly

based on either replacing or scaling metrics or parameters [44, 45, 46].

3.3.3 Allometric Scaling

Allometry describes the relationship between variables such as heart rate and body weight and can be applied to account, to some extent, for inter-species variability [47]. The allometric equation is a power law function given by,

$$y = k\, x^a \tag{3.30}$$

where x and y are the two variables, k the proportionality constant and a the scaling exponent. Studies have shown that the heart rate of an organism is approximately proportional to its body weight raised to the power of -0.25. Moreover, the proportionality constant is similar for organisms in the same taxonomic or functional group [48]. This leads to the following relationship between the heart rate of humans and mice,

$$\frac{y_{\text{Human}}}{BW_{\text{Human}}^{-0.25}} \approx \frac{y_{\text{Mouse}}}{BW_{\text{Mouse}}^{-0.25}}, \tag{3.31}$$

where the parameters y_i and BW_i denotes the heart rate and bodyweight of specie i, respectively. Heart rate in turn is correlated with oxygen consumption, and thus metabolic rate. This gives a rationale for using allometric scaling for inter-species translation of rate parameters such as the half-life of drugs. In pharmacometrics, allometric scaling is especially valuable when extrapolating PK parameters from animals to humans. Moreover, it has come to be the standard practice of pediatricians for dosing children, based on information regarding adults. Although attempts have been made to scale PD parameters as well, *e.g.*, growth rates, how well it works in practice is still up for debate. When considering inter-species translation, for example, the relationship is further complicated by the fact that human tumors are growing in a mouse microenvironment.

3.4 Machine Learning in Oncology

Machine Learning (ML) or artificial intelligence (AI) has emerged during the last decade as a very popular class of methods for analyzing large data sets that are used in numerous different fields. However, in essence, ML is parameter estimation. In this section, we discuss regression-based methods and how they are used in oncology to both identify important covariates and their influence on the EBEs from an NLME model [49].

Regression methods are often used to quantify the linear relationship between observed variables (the EBEs) and a set of explanatory variables (covariates). This is, of course, not a new problem and classical statistics has had the answer to how to perform this type of analysis since at least the start of the 19th century, namely linear regression (LR). The LR model describes the M observations of individual i, $y_i \in \mathbb{R}^M$, by a linear equation,

$$y_i = \beta \, g(x_i^T) + \epsilon_i \tag{3.32}$$

where $x_i \in \mathbb{R}^N$ and $\beta \in \mathbb{R}^{M \times N}$ are the $N - 1$ recorded covariates values of individual i and their influence on each observation, respectively, and $g : \mathbb{R}^N \mapsto \mathbb{R}^N$ is a potentially nonlinear function. Thus, the model is linear in the sense of the coefficients. We also note that the first element of x_i is set to 1 to account for a baseline coefficient and that ϵ is the same as in the PK/PD models, *i.e.,* the error between the model prediction and the observation. The β-vector can be estimated by considering the entire population of n individuals and solving,

$$\min_{\beta} \sum_{i=1}^{n} ||y_i - \beta \, g(x_i^T)||_2 \tag{3.33}$$

After calibrating the model, covariates whose influence cannot be estimated with high enough precision (often in terms of p-values) can be discarded and the model recalibration without those. Two similar regression methods that directly incorporate covariate selection in the model are LASSO (lowest absolute shrinkage and selection operator) and ridge regression [50]. Both methods include a penalty (or regularization) term to the minimization problem that is the lp-norm of the coefficients, *i.e.,* we solve,

$$\min_{\beta} \sum_{i=1}^{n} ||y_i - \beta\ g(x_i^T)||_2 + \lambda||\beta||_p^p, \tag{3.34}$$

where λ is a hyperparameter. $p = 1$ and $p = 2$ give LASSO and ridge regression, respectively.

4 Summary of Papers

4.1 Paper I

In this paper, published in BMC Cancer, the long-term radiation and radiosensitizer model first proposed by Cardilin *et al.* [51] was refined and further analyzed. The model is based on a system of ODEs that was calibrated with xenograft data from three studies, provided by Merck. The studies evaluated the efficacy of three separate radiosensitizers in combination with radiation treatment and the aim was to both fit the model to all data simultaneously and rank the radiosensitizers based on efficacy.

To accomplish this, both inter-study variability and inter-individual variability were quantified from the data. Furthermore, a Monte-Carlo-based simulation method for ranking compounds in terms of population TSE was developed. The method determines the necessary exposure pairs of radiation dose and radiosensitizer concentration that lead to tumor stasis for different population percentiles, *e.g.*, 50% or 95%. Two strengths of the simulation-based method are that it is suitable for complex models and directly incorporates the treatment schedule in the predictions. Therefore, it can potentially be used to select what compounds to proceed with in the subsequent drug development phase.

4.2 Paper II

The second paper was published in Cancer Chemotherapy and Pharmacology. In it, we investigate how well clinical oncology results can be predicted from a translated preclinical semi-mechanistic model. To accomplish this, volumetric xenograft tumor data was first searched for in the literature for combinations where published clinical data (RECIST response) also was available. Three

such combinations were found and the data was modeled using preclinical tumor growth inhibition models.

The preclinical models were translated by replacing mouse exposure with human exposure and allometric scaling of all PD rate parameters. Furthermore, we estimated optimal scaling parameters given the observed clinical data by solving an optimization problem. The aim was to find generally applicable scaling parameters that better describe the differences between human and xenograft mice models than what allometry proposes. Clinical predictions were performed using both scaling techniques and through a bootstrap procedure, 95% confidence intervals of these predictions were obtained.

4.3 Paper III

In the third paper, published in Clinical Pharmacology and Therapeutics: Pharmacometrics and Systems Pharmacology, a joint modeling approach for describing and predicting progression-free survival was constructed. The joint model consists of a tumor growth inhibition model in combination with a time-to-event model. In addition, we also used another parametric time-to-event model to account for uninformative dropout.

All models were calibrated with data coming from a clinical phase III study. The joint model in combination with the dropout model was able to describe the data well, and the predictive power of the models was tested by performing internal and external validations. For the internal validation, the model was recalibrated with truncated data and then used to predict the removed data. As an external validation, we predicted the progression-free survival of a missing treatment arm. In both cases, the model was shown to have good predictive capabilities. Regression and decision tree-based machine learning algorithms were also used to screen for important patient covariates influencing the EBEs.

4.4 Manuscript IV

In this paper, which is in manuscript form, analytical expressions for the probability that patients are categorized into each RECIST category were derived. The expressions were used to find an equation that predicts the necessary sample size of a clinical study to achieve a certain significance level and power. Thus, we were able to link the tumor growth inhibition model with power analysis, essentially creating a parametric model that can be used for sample size

calculations. Moreover, the probabilistic equations were also used to extend the classical TSE concept to align more with RECIST. By differentiating the probability functions with respect to the drug concentration, we also derived an equation that describes what concentration maximizes the probability that a patient is classified as stable disease.

4.5 Manuscript V

In the fifth manuscript, we investigated how the choice of population model affects the distribution of progression-free survival. We showed that there are qualitative differences in distribution between a commonly used preclinical model and a clinical model. Leveraging the information from this analysis should increase the translational potential of preclinical tumor models. Furthermore, we also considered how different assumptions of combination therapy efficacy lead to different distributions. Preclinical data was also modeled to investigate how well the independent-drug action hypothesis can be used in dynamical system modeling.

5 Discussion

In this concluding chapter of the thesis, we explore the primary contributions to the field and highlight key findings derived from the various appended papers. The aim is to put the results into a bigger perspective and to give an idea of how they can influence future research projects.

One of the biggest scientific contributions of this thesis is the continued development of the predictive parametric models meant to be used as an alternative to the nonparametric descriptive analysis methods that are often used today. The two most prominent examples come from Paper III and Manuscript IV, where parametric models were explored for both PFS survival curves and power analysis, respectively. Very important questions can be considered with the help of these models. For example, by modeling how the required sample size is affected by different parameters, *e.g.,* the drug dose, predictions can be made on how a change in treatment schedule changes the number of patients that has to be recruited for the study. Moreover, the PFS model was shown to have the potential to predict essential clinical metrics early in the trial period and to make predictions on the efficacy of new combination therapies. Further analysis should include validating the model on more data sets and refining the model with, *e.g.,* individual treatment schedules and dose reductions. The time-to-event models, which are essential for the model-based PFS approach, often require more mature data sets for accurate calibration, in comparison with the tumor growth inhibition models. Thus, an interesting avenue of continued research regarding this predictive approach could be to investigate if it is possible to leverage already existing time-to-event models for similar drugs and the same tumor type for making predictions on new drug candidates. The hope here is that since the drug-specific parameters are contained in the tumor growth inhibition model, the parameters of the time-to-event model reflect relationships of a more general nature, such as the correlation between tumor size and the probability of patient death.

The quantification of different levels of variability features heavily through-

out the thesis. In all papers and manuscripts, inter-individual variability is considered either from a theoretical standpoint or by quantifying it from experimental data. Furthermore, in all but Manuscript IV intra-individual variability, in terms of measurement error, is quantified. These two varieties of variability are frequently the minimum that must be considered when analyzing population data. However, as mentioned earlier, various other forms of variability must be taken into account in different situations.

Although inter-study variability and inter-occasional variability are mathematically the same thing, there is a point here to differentiate between them. Paper I is the only project where inter-study variability was considered and it turned out to be essential to quantify it to adequately model data from three different studies. This was surprising as the studies were carried out under the supervision of the same person, in the same lab, and with the same drug (although different formulations). The implication of this is that inter-study variability can be a major factor when comparing different studies to each other. It also helps to explain why researchers have found it hard to replicate results from studies carried out by other teams [52]. Due to data limitations, the influence of inter-study variability was not estimated in any other papers.

Inter-occasional variability is considered in Manuscript V to model the variability in replicates of PDXs. However, in practice, it was problematic to quantify it from the data because only one replicate was assigned to each treatment arm. It was therefore fixed to either zero or reasonable values taken from the literature for most parameters. This led to models that fit the data well and conformed with the expectations based on previous research by other scientists. This approach opens up further investigation into the individual effects of drugs when analyzing data from studies where multiple PDXs are created from several patient's tumors.

Preclinical studies have to be designed in such a way as to allow the inference of clinical conclusions. The inference step, *i.e.*, translation, is the final form of variability that we discuss. Model-based translational tools have been considered and advanced through the presented research, most notably in Paper II and Manuscript V. Here quantitative and qualitative inter-species differences were considered, respectively. The proof-of-concept for finding optimal translational scaling parameters outlined in Paper II is a good starting point for larger translational research efforts. However, the research presented in Manuscripts IV and V should be leveraged to refine the methodology and practical implementation of the algorithms. The analytical equations derived in Manuscript IV can be used to alleviate the need for some time-consuming numerical simulations, which is particularly useful if larger data sets are considered. Furthermore, considering clinical data with individual tumor size

time series would allow for the construction of PFS curves based only on target progression. This should allow a more accurate comparison to PFS curves estimated from mice. In addition, it enables the comparison of the entire PFS curve and not just the proportion of patients classified as progressive disease at week 8, *i.e.,* the PFS curve evaluated at week 8. The qualitative inter-species differences highlighted in Manuscript V could be essential to consider to allow for this analysis to be successful. A first step to make this possible would be to perform an extensive search of publically available preclinical and clinical data sets, in *e.g.,* ProjectDataSphere [53].

Bibliography

[1] Philip Gerlee and Torbjörn Lundh. *Scientific Models*. Springer International Publishing, Cham, 2016.

[2] Stephan Schmidt, Sarah Kim, Valvanera Vozmediano, Rodrigo Cristofoletti, Almut G. Winterstein, and Joshua D. Brown. Pharmacometrics, Physiologically Based Pharmacokinetics, Quantitative Systems Pharmacology—What's Next?—Joining Mechanistic and Epidemiological Approaches. *CPT: Pharmacometrics & Systems Pharmacology*, 8(6):352–355, 2019.

[3] Douglas Hanahan and Robert A. Weinberg. Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674, March 2011. Publisher: Elsevier.

[4] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, 2021.

[5] Natalia L. Komarova and C. Richard Boland. Calculated Treatment. *Nature*, 499(7458):291–292, 2013.

[6] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. Combinatorial Drug Therapy for Cancer in the Post-Genomic Era. *Nature Biotechnology*, 30(7):679–692, 2012.

[7] Adam C. Palmer and Peter K. Sorger. Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy. *Cell*, 171(7):1678–1691.e13, December 2017.

[8] T Philips, R Hoppe, and M Roach. *Leibel and Phillips Textbook of Radiation Oncology*. 3rd edition, 2010.

[9] L Harrison, M Chadha, R Hill, K Hu, and D Shasha. Impact of Tumor Hypoxia and Anemia on Radiation Therapy Outcomes. *The Oncologist*, 7(6):492–508, 2002.

[10] Reinhard Dummer, Paolo A. Ascierto, Helen J. Gogas, Ana Arance, Mario Mandala, Gabriella Liszkay, Claus Garbe, Dirk Schadendorf, Ivana Krajsova, Ralf Gutzmer, Vanna Chiarion-Sileni, Caroline Dutriaux, Jan Willem B. de Groot, Naoya Yamazaki, Carmen Loquai, Laure A. Moutouh-de Parseval, Michael D. Pickard, Victor Sandor, Caroline Robert, and Keith T. Flaherty. Encorafenib plus Binimetinib versus Vemurafenib or Encorafenib in Patients with BRAF-Mutant Melanoma (COLUMBUS): A Multicentre, Open-Label, Randomised Phase 3 Trial. *The Lancet Oncology*, 19(5):603–615, May 2018. Publisher: Elsevier.

[11] Jonathan B. Fitzgerald, Birgit Schoeberl, Ulrik B. Nielsen, and Peter K. Sorger. Systems Biology and Combination Therapy in the Quest for Clinical Efficacy. *Nature Chemical Biology*, 2(9):458–466, 2006.

[12] John W. Park, Robert S. Kerbel, Gary J. Kelloff, J. Carl Barrett, Bruce A. Chabner, David R. Parkinson, Jonathan Peck, Raymond W. Ruddon, Caroline C. Sigman, and Dennis J. Slamon. Rationale for Biomarkers and Surrogate End Points in Mechanism-Driven Oncology Drug Development. *Clinical Cancer Research*, 10(11):3885–3896, 2004.

[13] Christopher H. Lieu, Aik-Choon Tan, Stephen Leong, Jennifer R. Diamond, and S. Gail Eckhardt. From Bench to Bedside: Lessons Learned in Translating Preclinical Studies in Cancer Drug Development. *JNCI: Journal of the National Cancer Institute*, 105(19):1441–1456, 2013.

[14] Attila A. Seyhan. Lost in Translation: The Valley of Death across Preclinical and Clinical Divide – Identification of Problems and Overcoming Obstacles. *Translational Medicine Communications*, 4(1):18, 2019.

[15] S. P. Langdon, H. R. Hendriks, B. J. M. Braakhuis, G. Pratesi, D. P. Berger, Ø. Fodstad, H. H. Fiebig, and E. Boven. Preclinical Phase II Studies in Human Tumor Xenografts: A European Multicenter Follow-up Study. *Annals of Oncology*, 5(5):415–422, 1994.

[16] Harvey Wong, Edna F. Choo, Bruno Alicke, Xiao Ding, Hank La, Erin McNamara, Frank-Peter Theil, Jay Tibbitts, Lori S. Friedman, Cornelis E. C. A. Hop, and Stephen E. Gould. Antitumor Activity of Targeted and Cytotoxic Agents in Murine Subcutaneous Tumor Models Correlates with Clinical Response. *Clinical Cancer Research*, 18(14):3846–3855, 2012.

[17] Vahideh Vakil and Wade Trappe. Drug Combinations: Mathematical Modeling and Networking Methods. *Pharmaceutics*, 11(5):208, 2019. Publisher: Multidisciplinary Digital Publishing Institute.

[18] Jacob Leander, Joachim Almquist, Anna Johnning, Julia Larsson, and Mats Jirstrand. Nonlinear Mixed Effects Modeling of Deterministic and Stochastic Dynamical Systems in Wolfram Mathematica. *IFAC-PapersOnLine*, 54(7):409–414, 2021.

[19] Monolix 2021R2, Lixoft SAS, a Simulations Plus company.

[20] Johan Gabrielsson, Francis D. Gibbons, and Lambertus A. Peletier. Mixture Dynamics: Combination Therapy in Oncology. *European Journal of Pharmaceutical Sciences*, 88:132–146, 2016.

[21] Ming Zhao, Michelle A. Rudek, Aleksandr Mnasakanyan, Carol Hartke, Roberto Pili, and Sharyn D. Baker. A liquid chromatography/tandem mass spectrometry assay to quantitate MS-275 in human plasma. *Journal of pharmaceutical and biomedical analysis*, 43(2):784–787, January 2007.

[22] J. D. Wright, F. D. Boudinot, and M. R. Ujhelyi. Measurement and analysis of unbound drug concentrations. *Clinical Pharmacokinetics*, 30(6):445–462, June 1996.

[23] Triantafyllos Stylianopoulos and Rakesh K. Jain. Combining two strategies to improve perfusion and drug delivery in solid tumors. *Proceedings of the National Academy of Sciences*, 110(46):18632–18637, November 2013. Publisher: Proceedings of the National Academy of Sciences.

[24] Monica Simeoni, Paolo Magni, Cristiano Cammia, Giuseppe De Nicolao, Valter Croci, Enrico Pesenti, Massimiliano Germani, Italo Poggesi, and Maurizio Rocchetti. Predictive Pharmacokinetic-Pharmacodynamic Modeling of Tumor Growth Kinetics in Xenograft Models after Administration of Anticancer Agents. *Cancer Research*, 64(3):1094–1101, 2004. Publisher: American Association for Cancer Research Section: Experimental Therapeutics, Molecular Targets, and Chemical Biology.

[25] Tim Cardilin, Joachim Almquist, Mats Jirstrand, Alexandre Sostelly, Christiane Amendt, Samer El Bawab, and Johan Gabrielsson. Tumor Static Concentration Curves in Combination Therapy. *The AAPS Journal*, 19(2):456–467, 2017.

[26] R. K. Sachs, L. R. Hlatky, and P. Hahnfeldt. Simple ODE Models of Tumor Growth and Anti-Angiogenic or Radiation Treatment. *Mathematical and Computer Modelling*, 33(12):1297–1305, 2001.

[27] Yoichi Watanabe, Erik L. Dahlman, Kevin Z. Leder, and Susanta K. Hui. A Mathematical Model of Tumor Growth and Its Response to Single Irradiation. *Theoretical Biology and Medical Modelling*, 13(1):6, 2016.

[28] Mary Helen Barcellos-Hoff, Catherine Park, and Eric G. Wright. Radiation and the Microenvironment – Tumorigenesis and Therapy. *Nature Reviews Cancer*, 5(11):867–875, 2005.

[29] Kevin M. Foote, J. Willem M. Nissink, Thomas McGuire, Paul Turner, Sylvie Guichard, James W. T. Yates, Alan Lau, Kevin Blades, Dan Heathcote, Rajesh Odedra, Gary Wilkinson, Zena Wilson, Christine M. Wood, and Philip J. Jewsbury. Discovery and Characterization of AZD6738, a Potent Inhibitor of Ataxia Telangiectasia Mutated and Rad3 Related (ATR) Kinase with Application as an Anticancer Agent. *Journal of Medicinal Chemistry*, 61(22):9889–9907, November 2018. Publisher: American Chemical Society.

[30] Hao Wang, Xiaoyu Mu, Hua He, and Xiao-Dong Zhang. Cancer Radiosensitizers. *Trends in Pharmacological Sciences*, 39(1):24–48, January 2018.

[31] Sohyoung Her, David A. Jaffray, and Christine Allen. Gold nanoparticles for applications in cancer radiotherapy: Mechanisms and recent advancements. *Advanced Drug Delivery Reviews*, 109:84–101, January 2017.

[32] Yu Mi, Zhiying Shao, Johnny Vang, Orit Kaidar-Person, and Andrew Z. Wang. Application of nanotechnology to cancer radiotherapy. *Cancer Nanotechnology*, 7(1):11, December 2016.

[33] Ilker Etikan. The Kaplan Meier Estimate in Survival Analysis. *Biometrics & Biostatistics International Journal*, 5(2), February 2017.

[34] Jennifer Le-Rademacher and Xiaofei Wang. Time-To-Event Data: An Overview and Analysis Considerations. *Journal of Thoracic Oncology*, 16(7):1067–1074, July 2021.

[35] Patrick Schober and Thomas R. Vetter. Special Article: Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesthesia and Analgesia*, 127(3):792, September 2018. Publisher: Wolters Kluwer Health.

[36] Brandon George, Samantha Seals, and Inmaculada Aban. Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4):686–694, August 2014.

[37] D. Y. Lin. On the Breslow estimator. *Lifetime Data Analysis*, 13(4):471–480, December 2007.

[38] Almquist J, Leander J, and Jirstrand M. Using sensitivity equations for computing gradients of the FOCE and FOCEI approximations to the population likelihood. *Journal of pharmacokinetics and pharmacodynamics*, 42(3), June 2015. Publisher: J Pharmacokinet Pharmacodyn.

[39] Bernard Delyon, Marc Lavielle, and Eric Moulines. Convergence of a Stochastic Approximation Version of the EM Algorithm. *The Annals of Statistics*, 27(1):94–128, 1999. Publisher: Institute of Mathematical Statistics.

[40] Martin Bergstrand, Andrew C. Hooker, Johan E. Wallin, and Mats O. Karlsson. Prediction-Corrected Visual Predictive Checks for Diagnosing Nonlinear Mixed-Effects Models. *The AAPS Journal*, 13(2):143–151, 2011.

[41] Silvy Laporte-Simitsidis, Pascal Girard, Patrick Mismetti, Sylvie Chabaud, Hervé Decousus, and Jean-Pierre Boissel. Inter-Study Variability in Population Pharmacokinetic Meta-Analysis: When and How to Estimate It? *Journal of Pharmaceutical Sciences*, 89(2):155–167, 2000.

[42] Robert Andersson, Tobias Kroon, Joachim Almquist, Mats Jirstrand, Nicholas D. Oakes, Neil D. Evans, Michael J. Chappel, and Johan Gabrielsson. Modeling of Free Fatty Acid Dynamics: Insulin and Nicotinic Acid Resistance under Acute and Chronic Treatments. *Journal of Pharmacokinetics and Pharmacodynamics*, 44(3):203–222, 2017.

[43] Catherijne A J Knibbe, Klaas P Zuideveld, Leon P H J Aarts, Paul F M Kuks, and Meindert Danhof. Allometric Relationships between the Pharmacokinetics of Propofol in Rats, Children and Adults. *British Journal of Clinical Pharmacology*, 59(6):705–711, June 2005.

[44] Alexander B. Herman, Van M. Savage, and Geoffrey B. West. A Quantitative Theory of Solid Tumor Growth, Metabolic Rate and Vascularization. *PLOS ONE*, 6(9):e22973, 2011.

[45] Dean C. Bottino, Mayankbhai Patel, Ekta Kadakia, Jilai Zhou, Chirag Patel, Rachel Neuwirth, Natasha Iartchouk, Rachael Brake, Karthik Venkatakrishnan, and Arijit Chakravarty. Dose Optimization for Anticancer Drug Combinations: Maximizing Therapeutic Index via Clinical Exposure-Toxicity/Preclinical Exposure-Efficacy Modeling. *Clinical Cancer Research*, 25(22):6633–6643, 2019. Publisher: American Association for Cancer Research Section: Precision Medicine and Imaging.

[46] Andy Zx Zhu. Quantitative translational modeling to facilitate preclinical to clinical efficacy & toxicity translation in oncology. *Future science OA*, 4(5):FSO306, 2018.

[47] Geoffrey B. West, William H. Woodruff, and James H. Brown. Allometric Scaling of Metabolic Rate from Molecules and Mitochondria to Cells and Mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 1:2473–2478, 2002.

[48] Geoffrey B. West, James H. Brown, and Brian J. Enquist. A General Model for the Origin of Allometric Scaling Laws in Biology. *Science*, 276(5309):122–126, 1997. Publisher: American Association for the Advancement of Science Section: Report.

[49] Nadia Terranova, Jonathan French, Haiqing Dai, Matthew Wiens, Akash Khandelwal, Ana Ruiz-Garcia, Juliane Manitz, Anja von Heydebreck, Mary Ruisi, Kevin Chin, Pascal Girard, and Karthik Venkatakrishnan. Pharmacometric modeling and machine learning analyses of prognostic and predictive factors in the JAVELIN Gastric 100 phase III trial of avelumab. *CPT: Pharmacometrics & Systems Pharmacology*, 11(3):333–347, 2022.

[50] Trevor Hastei, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2th edition, 2017.

[51] Tim Cardilin, Joachim Almquist, Mats Jirstrand, Astrid Zimmermann, Floriane Lignet, Samer El Bawab, and Johan Gabrielsson. Modeling Long-Term Tumor Growth and Kill after Combinations of Radiation and Radiosensitizing Agents. *Cancer Chemotherapy and Pharmacology*, 83(6):1159–1173, 2019.

[52] John P. A. Ioannidis. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA*, 294(2):218–228, 2005.

[53] Project Data Sphere, 2022.