# Modelling choices in Pharmacokinetic and Pharmacodynamic models


Rebecca Rumney
Wolfson College
University of Oxford

A thesis submitted for the degree of
*Master of Science by Research*

Michealmas 2023

# Acknowledgements

I would like to thank my academic Supervisors, Professor Ruth Baker, Professor David Gavaghan, and Dr Ben Lambert for all their tremendous help over the years, and my Industrial supervisors, Antje-Christine Walz, and Ken Wang for all that they have taught me.

I would also like to thank the entirety of the WCMB for making me feel welcome even through a pandemic and providing a decent amount of cake.

I would especially like to thank Alex and my mum for the emotional and practical support that they have given. I would also like to thank the rest of my and Alex's family.

I would like to thank the 2019 SABS cohort, the PINTS team, and the friends that I have made through this journey.

And finally I would like to thank the cats who kept me company while writing.

# Abstract

Pharmacokinetic and pharmacodynamic (PKPD) models can be used to predict the benefits or toxicity from drugs, such as the neutropaenia, anaemia and thrombocytopaenia caused by chemo-therapeutic drugs. In this thesis, I will be exploring the methodology used to make modelling decisions in these PKPD models, and apply them to a commonly used model by Friberg *et al.* [17]. I will use these methods to compare different ways to model the noise observed in the data and to build a population-based model. Using Bayesian methods to determine model parameters, profile likelihoods to ensure identifiability, and Widely Applicable Information Criteria (WAIC) for model selection, I found that only multiplicative or constant Gaussian noise were fully identifiable and constant Gaussian noise had the greatest out-of-sample predictive power. These methods were also utilised to build a mixed-effects population model and I found that a single mixed-effects parameter had better out-of-sample predictive power, and including further parameters affected convergence of the Monte Carlo samplers in Bayesian inference. I also developed a more computationally efficient method for acquiring Profile likelihoods to determine practical identifiability. I propose that the methodology outlined in this thesis for making modelling decisions should be standard for a general PKPD case. However, there may need to be further work done to ensure it is suitable for all modelling needs.

# Contents

# Chapter 1

# Introduction

A common adverse effect of many drugs used in chemotherapy is neutropaenia, dangerously low levels of neutrophils, ∼45.1% of patients) which can lead to severe or fatal infection [26, 46]. Other common side-effects are, anaemia (∼22.4% of patients), which leads to severe fatigue, and thrombocytopaenia (∼9.7% of patients), which reduces blood clotting and causes bleeding complications [21, 63, 65]. The incidence of these adverse effects have also increased over the last few decades, due to new, more effective chemotherapy treatments also having greater myelotoxic effects [47].

When these complications are observed in patients, the most common response is to delay treatment or reduce the dose until the blood cell count recovers. However, this leads to a reduced chance of becoming disease-free or surviving [13]. Alternatively, growth factors are used in approximately 50% of patients to stimulate the regrowth of blood cells [46], and risk factors, such as baseline level of blood cell count, age or disease specific factors, can be assessed before chemotherapy begins to ensure that these treatments are administered in a preventative manner [67, 53]. However, these treatments are expensive and significantly reduce the quality of life of the individual [10].

These complications can be partially avoided by utilising pharmacokinetic and pharmacodynamic (PKPD) models. These mathematical models simulate the relationship between a dosing regimen and an observed effect in two steps. First, a pharmacokinetic (PK) model predicts the time course of drug concentration in a relevant biological fluid (such as blood plasma) after a given dose. The second step is to relate this drug concentration to an outcome. Pharmacodynamic (PD) models achieve this by translating drug concentration (or PK predictions of drug concentrations) into a measurable PD effect, such as cell count. This modelling may be done at the drug development stage, to predict the occurrence of myelotoxicity in a population for a certain dosing regimen. By predicting outcomes from many different

adjustments to the dosing regimen, the dose can be optimised to ensure maximum efficacy while avoiding myelotoxicity, or to compare the toxicity of different anti-cancer drugs. There is also the potential for PKPD models to be utilised in personalising treatments by predicting the blood cell count over time given the varying risk factors of an individual. This prediction is updated regularly throughout treatment, given the patient's response to the chemotherapy, and provides insight on how to adapt treatment over time.

PKPD models are becoming more utilised throughout the drug development pipeline, especially as a way to reduce the cost of drug development and better predict earlier in the pipeline, which drugs will fail to get to market. The U.S. Food and Drug Administration and the European Medicines Agency have both encouraged the use of modelling in drug development [49, 24]. However, there is currently a limit in the extent to which PKPD models can be used: "The shortage of analysts with training and experience in this area is one major impediment for the application of the methodology." [1] Very large and costly clinical trials are still required in most cases to show the efficacy and safety of the drug.

These limitations mainly stem from the complexity and variety in the analysis of these models. There is a wide selection of methods that a modeller might choose from, and many of these have not been built specifically for PKPD models. This means a modeller needs to be trained to identify which methods to use and to adapt those methods for their particular use-case. The chosen methods will also differ from one modeller to another, making it difficult for others to compare studies, reducing confidence in the results. I aim to contribute to solving this problem by building towards a standardised methodology in PKPD model analysis. This standardisation will lead to clinicians requiring less experience to implement and understand these methods, and for PKPD model analysis to be partially automated, reducing the burden on specialists. This approach would allow easier comparison between competing research approaches and increase confidence in the predictions of PKPD models. This standardisation and automation may also allow modelling to be utilised in the personalisation of medication by medical practitioners.

In this thesis I will explore a PKPD model of chemotherapy-induced myelotoxicity developed by Friberg *et. al.* [17]. I am using this case example to build a standard approach to PKPD model analysis including how choices in the modelling process can be made. In Chapter 2, I will start with how PKPD models are built and how initial modelling choices are made given the known mechanisms involved and the identifiable behaviours seen in data. In most cases, however the model and parameters that best

describe the true system are not known, but the resulting time-profile is known (at certain time points, of a sample of individuals and within some error). This describes the inverse problem: given a data-set, which model and set of parameters can best describe the observations. In the rest of the thesis I will be primarily exploring methods that combat this problem, utilising both simulated and experimental data which are described in Chapter 4. I have identified that most methods for analysis of the inverse problem attempt to answer one of these three questions:

- Given some data, what parameter values of a model should be used to best describe the data?
- How do we ensure that these parameter values can be reasonably identified?
- Given a selection of models, what is the best model to use?

The first question is answered using parameter inference which optimises an objective function to fit the model output to the data. Parameter inference is commonly performed in a range of PKPD studies, where comparisons to data are made, such as cancer modelling [15], cardiac modelling [32] and toxicity modelling [51]. I use two inference techniques, maximum likelihood estimation in Chapter 5 and Bayesian inference in Chapter 7 to determine this. The second question is answered using identifiability analysis, which is concerned with the uniqueness of the model parameters, and can be defined in two ways. Structural identifiability analysis assumes that there is infinite noiseless data, and determines, for any model output, whether there is a unique vector of parameter values that produces that output. I explore this in Chapter 3. Practical identifiability analysis is linked to a particular dataset and determines whether the global optimum, found during parameter inference, is significantly more optimal than other local optima. This is explored in Chapter 6. These analyses are less common in the literature but gaining prevalence in new studies [61] and being used to analyse existing, commonly-used models [29, 30]. I will answer the third question in Chapter 8 where I will perform methods for comparing models that penalise over-fitting. This is generally performed where the researcher has proposed multiple potential models or they are adapting a pre-existing model [23, 5]. This all builds towards using PKPD to predict outcomes as explored in Chapter 9.

# Chapter 2

# Forward models

To model the relationship between a dosing regimen and an observed effect, two steps need to be taken. First, a PK model needs to be implemented to predict the time course of drug concentration in a relevant biological fluid (such as blood plasma) after a given dose. The second step is to relate this drug concentration to an outcome. PD models achieve this by translating drug concentration or PK predictions of drug concentrations) into a measurable PD effect, such as cell count. In this section, I will review the existing forward models from the literature, that are relevant to the myelotoxicity problem.

## 2.1 The pharmacokinetic model

The empirical PK model that is most commonly used separates the body into a number of hypothetical compartments, see Figure 2.1. There is the central compartment, which usually encompasses blood plasma, and a number of peripheral compartments. The drug in the central compartment can be transferred to the peripheral compartments and vice versa, and it is also cleared from the central compartment. Additionally, depending on the route of dose administration, there may be a dosing compartment. The $n$-compartment PK model is given by the following system of ordinary differential equations (ODEs) [27],

$$\begin{aligned}
\frac{\mathrm{d}A_d}{\mathrm{d}t} &= -\, q_d \left( \frac{A_d\,(t)}{V_d} \right), \\
\frac{\mathrm{d}A_c}{\mathrm{d}t} &= q_d \left( \frac{A_d\,(t)}{V_d} \right) - q_{\mathrm{cl}} \left( \frac{A_c\,(t)}{V_c} \right) - \sum_{i=1}^{n-1} q_i \left( \frac{A_i\,(t)}{V_i}, \frac{A_c\,(t)}{V_c} \right), \\
\frac{\mathrm{d}A_i}{\mathrm{d}t} &= q_i \left( \frac{A_i\,(t)}{V_i}, \frac{A_c\,(t)}{V_c} \right), \qquad 1 \leq i < n,
\end{aligned}$$

where $A_d$, $A_c$ and $A_i$ are the masses of drug in the dosing, central and $i$-th peripheral compartments, respectively, $V_d$, $V_c$ and $V_i$ are the volumes of the dosing, central and $i$-th peripheral compartments, respectively, $q_i$ is a function describing the flux of the drug into the $i$-th peripheral compartment, $q_d$ is a non-negative function describing the flux of the drug into the central compartment from the dosing compartment, and $q_{\text{cl}}$ is a non-negative function describing the loss of the drug from the central compartment through clearance. The peripheral compartments have initial conditions $A_i(0) = 0$, the dosing compartment has initial condition $A_d(0) = d$, the dose amount, or $A_d(0) = 0$, depending on the method of dose administration. Similarly, the central compartment has initial condition $A_c(0) = 0$ or $A_c(0) = d$ depending on the method of dose administration [41, 42, 44].

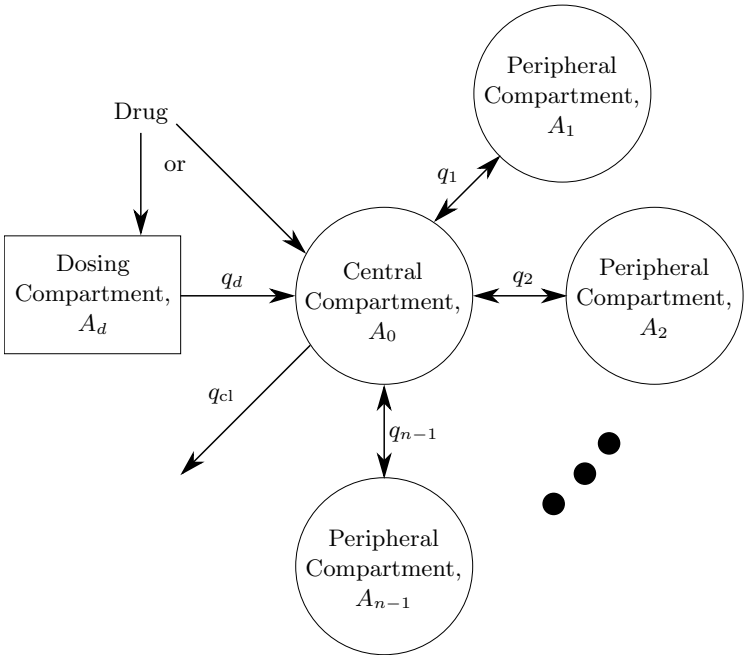



Figure 2.1: Diagram of the general $n$-compartment PK model. The drug enters the body either through a dosing compartment, or straight into the blood stream, represented as the central compartment. It then moves in and out of the different peripheral compartments, before being cleared from the central compartment.

This empirical PK modelling requires three decisions to be made:

- whether a dosing compartment is required;
- the number of compartments to model;
- the form the rates of transfer and clearance should take;

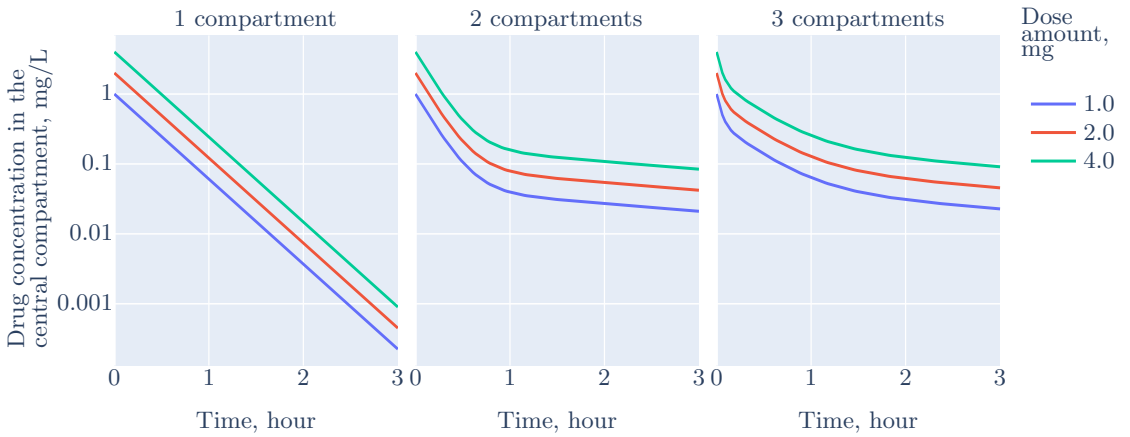



Figure 2.2: Simulations of the compartmental PK model with one-compartment (left) using Equation 2.1, two-compartment (middle) using Equations 2.2-2.3, and three-compartment (right) using Equations 2.4-2.6. Transfer rates between compartments are linear and no dosing compartment is included. Parameters, when applicable to the model, are $V_c = 1$, $K_{\text{cl}} = 2.8$, $V_1 = 4.8$, $K_1 = 2.3$, $V_2 = 0.9$ and $K_2 = 9.2$. $y$-axis is on a log-scale.

All of which affect the resulting concentration-time curves. To choose whether to include a dosing compartment simply depends on the form of drug administration that is being modelled. For example, if the drug is administered intravenously (IV), then no dosing compartment is required, but if the drug is administered subcutaneously then the dosing compartment is required. The number of compartments and the form of the transfer rate are decided based on the expected shape of the concentration-time curve plotted on a log scale (see Figure 2.2). If one compartment is used, this plot is a straight line (due to there being a single exponential term in the solution). If two compartments are used then two linear sections can be seen in the plot (due to two exponential terms in the solution). And for $n$-compartments, there can be up to $n$ linear sections in the plot. These linear sections can be hard to detect for large $n$, but in these cases a simpler model is usually sufficient to capture the same dynamics. For deciding the form of the rates used, the concentration-time profiles for multiple doses need to be plotted. The simplest form these rates could take is a linear form,

i.e.

$$
\begin{aligned}
q_d\left(\frac{A_d}{V_d}\right) &= K_d\frac{A_d}{V_d},\\
q_{\text{cl}}\left(\frac{A_c}{V_c}\right) &= K_{\text{cl}}\frac{A_c}{V_c},\\
q_i\left(\frac{A_i}{V_i},\frac{A_c}{V_c}\right) &= K_i\left(\frac{A_c}{V_c}-\frac{A_i}{V_i}\right),
\end{aligned}
$$

where $K_d > 0$ is the rate of transfer from the dosing compartment to the central compartment, $K_{\text{cl}} > 0$ is the rate of clearance from the central compartment and $K_i > 0$ is the rate of transfer between the central compartment and the $i$-th peripheral compartment.

Thus for the linear rates the one, two and three-compartment PK models, assuming no dosing compartment, would produce the following systems of ODEs:

$$\frac{\mathrm{d}A_c}{\mathrm{d}t} = -K_{\text{cl}}\frac{A_c}{V_c}, \tag{2.1}$$

for the one-compartment model;

$$\frac{\mathrm{d}A_c}{\mathrm{d}t} = -K_{\text{cl}}\frac{A_c}{V_c} - K_1\left(\frac{A_c}{V_c}-\frac{A_1}{V_1}\right), \tag{2.2}$$

$$\frac{\mathrm{d}A_1}{\mathrm{d}t} = K_1\left(\frac{A_c}{V_c}-\frac{A_1}{V_1}\right), \tag{2.3}$$

for the two-compartment model; and

$$\frac{\mathrm{d}A_c}{\mathrm{d}t} = -K_{\text{cl}}\frac{A_c}{V_c} - K_1\left(\frac{A_c}{V_c}-\frac{A_1}{V_1}\right) - K_2\left(\frac{A_c}{V_c}-\frac{A_2}{V_2}\right), \tag{2.4}$$

$$\frac{\mathrm{d}A_1}{\mathrm{d}t} = K_1\left(\frac{A_c}{V_c}-\frac{A_1}{V_1}\right), \tag{2.5}$$

$$\frac{\mathrm{d}A_2}{\mathrm{d}t} = K_2\left(\frac{A_c}{V_c}-\frac{A_2}{V_2}\right), \tag{2.6}$$

for the three compartment model. The initial conditions for these equations are $A_1(0) = A_2(0) = 0$, and $A_c(0) = d$

If this is the case, then the concentration-time profiles of the different doses would be parallel. However, if there are non-parallel sections between the doses then at least one of these rates must be non-linear. This can also be checked by plotting the dose-normalised (drug concentration/dose amount) curve, where for a linear system, the curve for each dose would perfectly overlap.

In practice, the model choice comes from the data that is collected. For example, the data shown in Figure 2.3, from IV bolus of Docetaxel in rats, shows clearly at least

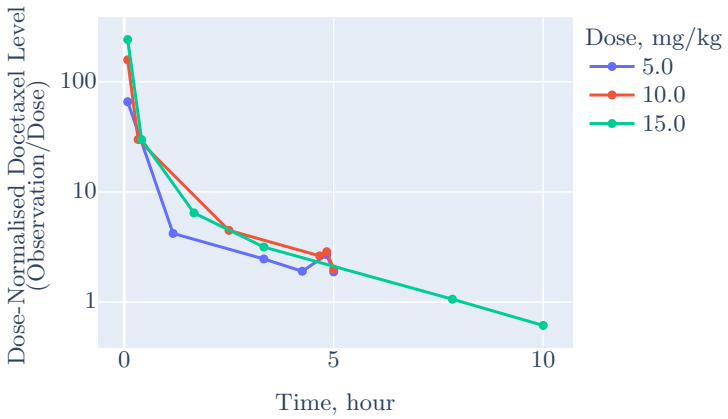



Figure 2.3: Experimental data of rats given different levels of Docetaxel. Observations (Docetaxel concentration, ng/mL) have been normalised by dose amount and then averaged over the rats in each dose group. Y-axis is on a log-scale. This data has been provided by Roche, see Chapter 4.

two linear sections and the dose-normalised curves overlap for the majority of points. Thus the most appropriate model for this drug would be a linear two-compartment model with no dosing compartment.

It is also possible to have a fully physiologically-based pharmacokinetic (PBPK) model where each organ and tissue is modelled as a compartment, all connected via the blood compartment [50]. This model can be more informative of the biological mechanics occurring with PK. However, it leads to a complicated system of non-linear ODEs with many parameters that are then generally difficult to estimate from data.

## 2.2 The pharmacodynamic model

To translate the concentration of drug in the central compartment, $C_0(t) = V_c^{-1} A_c(t)$, into a direct drug effect, $E_{\text{drug}}$, the PD relation is commonly assumed to be either a linear relation,

$$E_{\text{drug}} = SC_0(t), \tag{2.7}$$

where $S > 0$ is the slope of the linear effect; or a non-linear $E_{\max}$ relation of the form

$$E_{\text{drug}}(t) = \frac{E_{\max} C_0(t)}{C_0(t) + EC_{50}}, \tag{2.8}$$

where $E_{\max} > 0$ is the maximum response and $EC_{50}$ is the concentration of drug at which the effect is at 50% of $E_{\max}$. The differences between these relationships are shown in Figure 2.4.

Unfortunately, in some cases, the direct drug effect is not observable and an indirect response model, which relates drug effect to drug response, is required. This is the case for myelotoxicity where the drug effect, the level of inhibition of proliferation, is not directly measurable but the response, the number of blood cells in circulation, is measurable. The model presented by Dayneka *et al.* [11] captures this type of indirect response. In this paper, the authors present four different models of action a drug could take on the response variable, $R$, the variable of interest to the researcher. In the myelotoxicity case, $R$ represents the concentration of proliferating blood cells in bone marrow. These models are of a general biological system where the response variable is produced at a constant rate of $K_{\text{in}} > 0$ and eliminated at a rate $K_{\text{out}} > 0$. The drug could affect either the production or the loss of $R$; and it is either inhibitory or stimulatory. For myelotoxicity, the production of blood cells is inhibited by the drug, and so the ODE for the response variable is

$$\frac{\mathrm{d}R}{\mathrm{d}t} = K_{\text{in}} \left(1 - E_{\text{drug}}\left(t\right)\right) - K_{\text{out}} R, \tag{2.9}$$

where $E_{\text{drug}}\left(t\right) < 1$ is the inhibitory drug effect as defined in Equation 2.7 or 2.8. The initial condition, $R\left(0\right) = R_0$, is such that the system begins at the drug-free steady state, i.e. $K_{\text{in}} = K_{\text{out}} R_0$. Figure 2.5 shows a diagram of the model and the output of the model.

However, there are some features seen in myelotoxicity data that this general model cannot capture, such as a delay between drug administration and the observed response of a drop in blood cell concentrations, and a peak of blood cell count during recovery before returning to the base level, as can be seen in Figure 2.6.

The response delay comes from the conversion of proliferating cells into circulating cells because it takes time for the cells to mature and enter circulation. As the proliferating cells are affected by the drug but the circulating cells are observed, this delay is seen in the data. Measuring the response in bone marrow directly in patients is not possible, thus a more specific model is required to capture this delay. One such model has been presented by Zamboni *et al.* [66] who extended the indirect response model from one to three compartments (see Figure 2.7). One compartment represents the proliferating cells, $P$, which the drug affects the production of; another compartment represents the measurable cell count in circulation around the body, $R$; in between these is a transit compartment, $T$, to produce the time delay. This leads

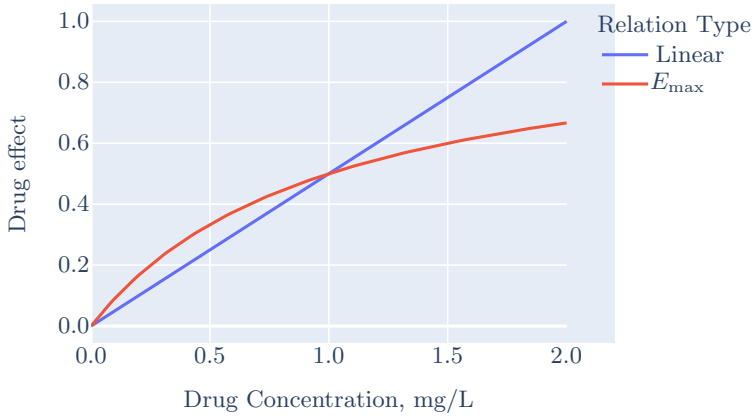



Figure 2.4: Comparison of linear (Equation 2.7) and non-linear (Equation 2.8) PD drug effect at different concentrations of drug. Parameters used in this diagram are $S = 0.5$ and $E_{\max} = EC_{50} = 1$

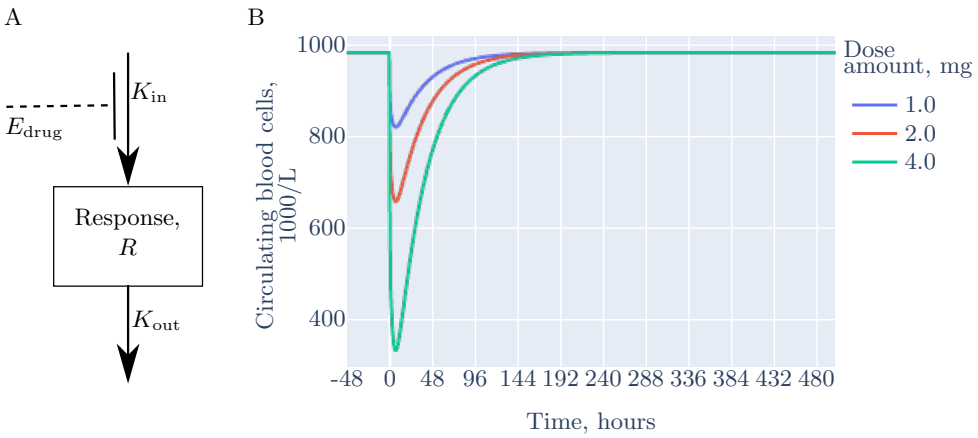



Figure 2.5: A) General indirect response model, where production of the response variable is inhibited by the drug. B) The response variable, $R$, over time from a simulation of this indirect response model (Equation 2.9, with parameters $R_0 = 983.1$ $10^3/\mu$L and $K_{\text{out}} = 0.03$ hr$^{-1}$). The drug effect is linear (Equation 2.7, with $S = 20$ L/mg) and the two-compartment PK model is used to generate the drug concentration over time (Equations 2.2-2.3, with parameters $V_c = 1$, $K_{cl} = 2.8$, $V_1 = 4.8$ and $K_1 = 2.3$).

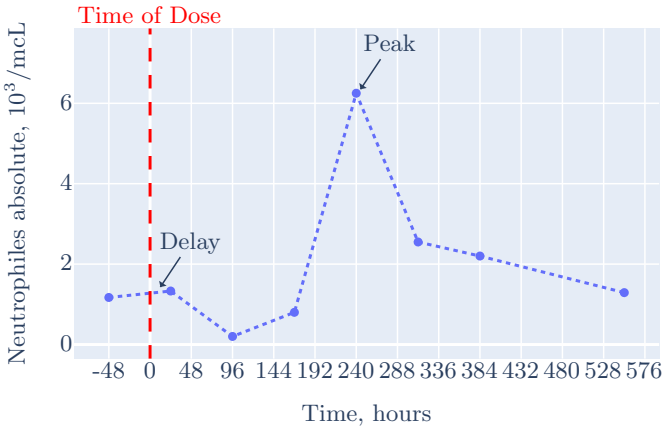



Figure 2.6: Time course of neutrophil concentration in the blood of a rat that had a single dose of 15mg/kg Docetaxel. It can be seen that there is a delay between the time of dose and the effect, as well as a large peak in neutrophil concentration above base level as the patient recovers. This data has been provided by Roche, see Chapter 4.

to a system of three ODEs:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = K_{\text{in}}\left(1 - E_{\text{drug}}\left(t\right)\right) - K_{\text{tr}}P, \tag{2.10}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = K_{\text{tr}}P - K_{\text{tr}}T, \tag{2.11}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = K_{\text{tr}}T - K_{\text{out}}R, \tag{2.12}$$

where $K_{\text{tr}} > 0$ is the rate of transfer between the compartments. The initial conditions are such that the system is initially at equilibrium, i.e. $P\left(0\right) = K_{\text{in}}/K_{\text{tr}}$, $T\left(0\right) = K_{\text{in}}/K_{\text{tr}}$, and $C\left(0\right) = K_{\text{in}}/K_{\text{out}} = R_0$.

The delay in response can be altered by adjusting the number of transit compartments used such that the equations become

$$\frac{\mathrm{d}P}{\mathrm{d}t} = K_{\text{in}}\left(1 - E_{\text{drug}}\left(t\right)\right) - K_{\text{tr}}P, \tag{2.13}$$

$$\frac{\mathrm{d}T_1}{\mathrm{d}t} = K_{\text{tr}}P - K_{\text{tr}}T_1, \tag{2.14}$$

$$\frac{\mathrm{d}T_i}{\mathrm{d}t} = K_{\text{tr}}T_{i-1} - K_{\text{tr}}T_i, \qquad 2 \leq i \leq n_T \tag{2.15}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = K_{\text{tr}}T_{n_T} - K_{\text{out}}R, \tag{2.16}$$

for $n_T$ transit compartments, $T_i$. Even when the mean transit time,

$$MTT = \left(n_T + 1\right)/K_{\text{tr}},$$

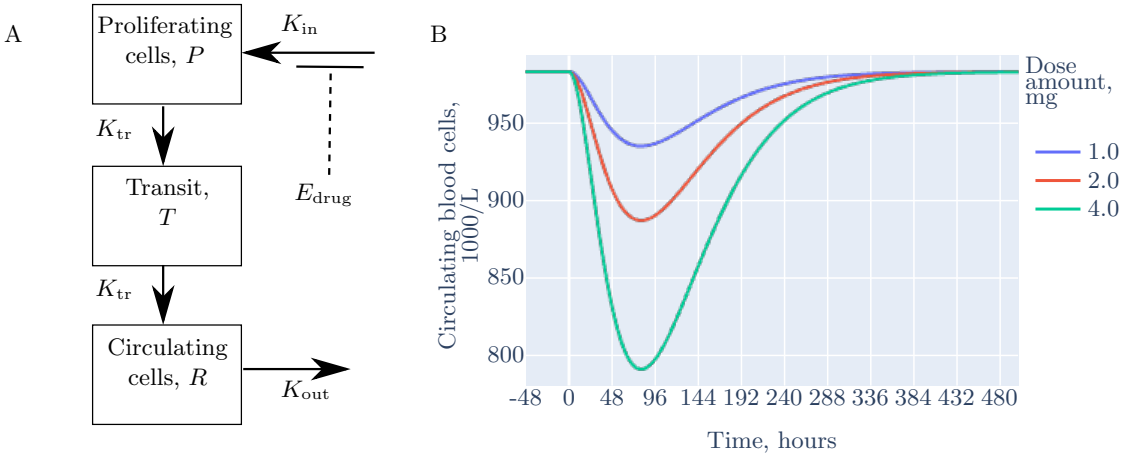



Figure 2.7: A) The Zamboni model of myelotoxicity, where proliferating cells enter a transit stage before entering into the circulation stage, and production of the proliferating cells is inhibited by the drug. B) The circulating blood cell concentration over time from a simulation of the Zamboni model (Equations 2.10 - 2.12). The PK model, drug effect model and parameters used for this graph were the same as in Figure 2.5, with $MTT = 2/K_{\text{tr}} = 85.26$ hr.

is kept constant and $n_T$ is altered the resulting plots show significant differences, as seen in Figure 2.8.

In the investigation by Friberg *et al.* [17], it was found that the inclusion of three transit compartments best captured the response delay in myelotoxicity. Another characteristic, often seen in the data, that Friberg *et al.* sought to capture in their model was a rebound in blood cells causing a peak above the baseline blood cell count. The above models cannot capture this characteristic as after the initial drop in blood cells count, the variable $R$ monotonically increases until at steady state. A potential biological explanation of this rebound effect is the presence of a negative feedback loop, where the system adjusts proliferation given the current level of circulating blood cells [52]. This feedback, coupled with the delay in response, causes the blood cell concentration to oscillate towards the steady state, instead of a monotonic convergence, producing additional peaks and troughs. Friberg *et al.* took the above model and adapted it to include negative feedback [17] (see Figure 2.9). The model

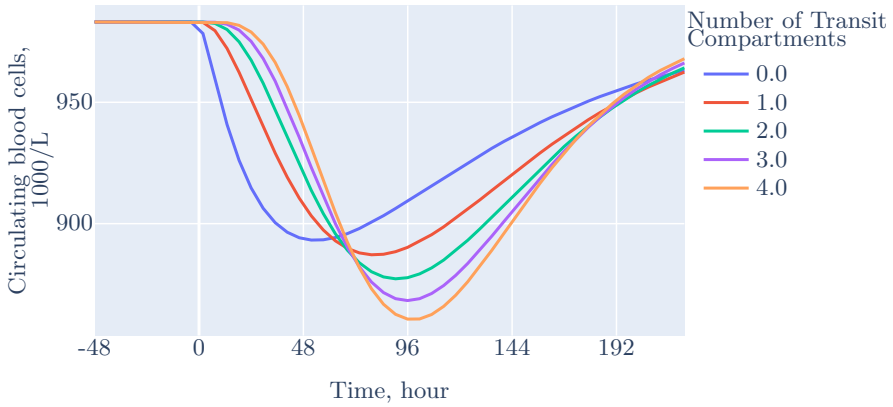



Figure 2.8: The circulating blood cell concentration over time from simulations of the Zamboni model (Equations 2.13-2.16) with varying number of transit compartments, $n_T$, and constant $MTT$. The PK model, drug effect model and parameters used for this graph were the same as in Figure 2.5, with $MTT = (n_T + 1)/K_{\text{tr}} = 85.26$ hr

is specified as the following system of ODEs:

$$\frac{\mathrm{d}P}{\mathrm{d}t} = K_{\text{prol}}\left(1 - E_{\text{drug}}(t)\right)\left(\frac{R_0}{R}\right)^{\gamma} P - K_{\text{tr}}P, \tag{2.17}$$

$$\frac{\mathrm{d}T_1}{\mathrm{d}t} = K_{\text{tr}}P - K_{\text{tr}}T_1, \tag{2.18}$$

$$\frac{\mathrm{d}T_2}{\mathrm{d}t} = K_{\text{tr}}T_1 - K_{\text{tr}}T_2, \tag{2.19}$$

$$\frac{\mathrm{d}T_3}{\mathrm{d}t} = K_{\text{tr}}T_2 - K_{\text{tr}}T_3, \tag{2.20}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = K_{\text{tr}}T_3 - K_{\text{out}}R, \tag{2.21}$$

where $R_0 > 0$ is the base level of circulating blood cells, $\left(\frac{R_0}{R}\right)^{\gamma}$ with $\gamma > 0$ is the negative feedback term, and $P(0) = T_i(0) = R(0) = R_0$ for $i \in \{1, 2, 3\}$.

As this negative feedback acts directly on the proliferation, another change made to the Zamboni model is to move from a constant influx, $K_{\text{in}}$ of proliferating cells (representing general maintenance of cell population from proliferation, stem cells and quiescent cells), to a proliferation term, $K_{\text{prol}}P$. This specific form of feedback loop can pose difficulties as $R \to 0$ which leads the rate of proliferation tending to infinity. According to the friberg paper, the feedback loop takes this form "because it is known that the proliferation rate can be affected by endogenous growth factors and cytokines and that circulating neutrophil counts and the growth factor G-CSF levels are inversely related." [17] In the data we are using (as discussed in Chapter

4), the levels of blood cells remains mostly within 250 to 1750 $10^3/\mu$L, and thus I do not expect $R$ to get too close to 0.

To ensure a non-trivial steady state in the absence of drug, $K_{\text{prol}} = K_{\text{tr}}$. Friberg *et al.* also set $K_{\text{out}} = K_{\text{tr}}$ which ensures the steady state for all variables to be $R_0$ and reduces the model parameters.

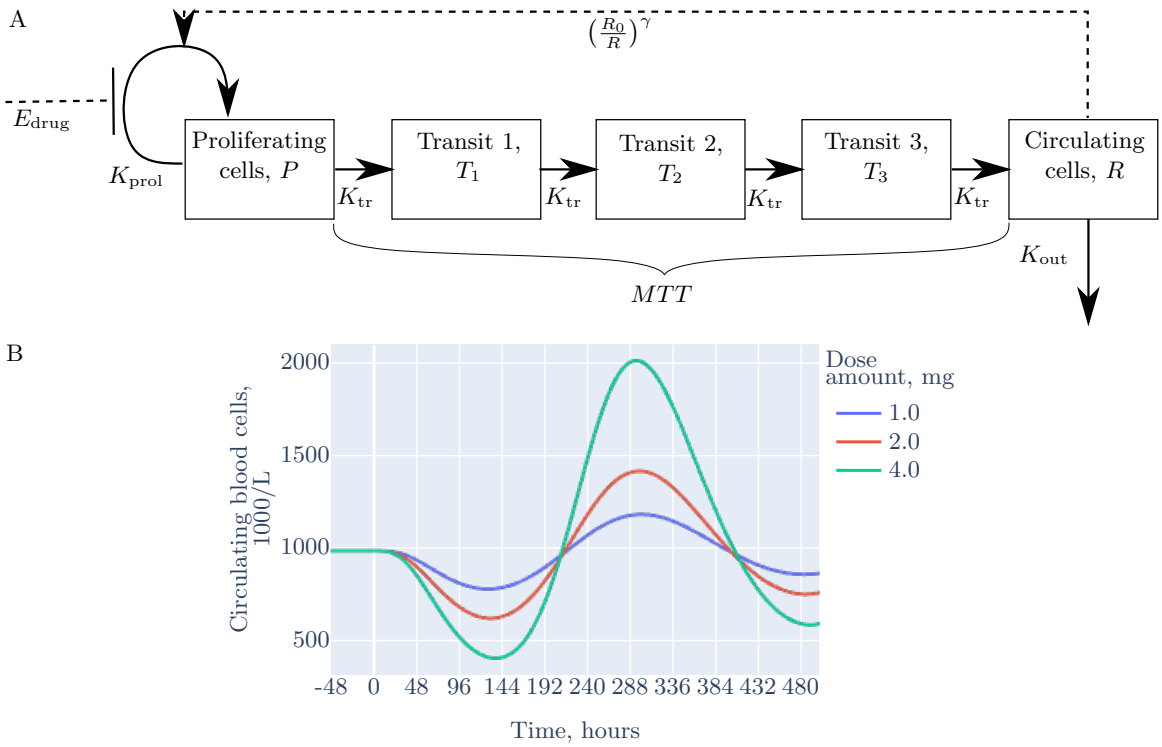



Figure 2.9: A) The Friberg model of myelotoxicity, where proliferating cells go through three transit stages before entering into circulation, proliferation is inhibited by the drug and there is negative feedback from circulating cells. B) The circulating blood cell concentration over time from a simulation of the Friberg model (Equations 2.17-2.21). The PK model, drug effect model and parameters used for this graph were the same as in Figure 2.5, with $MTT = 4/K_{\text{tr}} = 85.26$ hr and $\gamma = 0.44$

## 2.3 The noise model

In experiments, due to inaccuracies in measurement and stochastic fluctuations in drug concentration and blood cell counts, the observation of the variable will differ from the simulated deterministic model. This noise, $\nu(t_j)$, can be incorporated into

a model by letting

$$x_j^{\text{obs}} = x^{\text{sim}}(t_j) + \nu(t_j), \tag{2.22}$$

where $x_j^{\text{obs}} \in X^{\text{obs}}$ is the observation at time $t_j$ and $x^{\text{sim}}$ is the output of the forward model. This noise is assumed to take one of multiple potential forms. The most commonly used is constant Gaussian noise where

$$\nu(t_j) \overset{i.i.d.}{\sim} N\left(0, \sigma_c^2\right), \tag{2.23}$$

for $\sigma_c \in \mathbb{R}^+$. An alternative noise is multiplicative Gaussian noise,

$$\nu(t_j) = \sigma_m x^{\text{sim}}(t_j)^\eta \epsilon_j, \tag{2.24}$$

for some $\sigma_m, \eta \in \mathbb{R}^+$ with $\epsilon_j \overset{i.i.d.}{\sim} N(0,1)$. For some systems, neither of these noise models fully capture the noise seen, instead a combination of these models is used by setting

$$\nu(t_j) = \left(\sigma_c + \sigma_m x^{\text{sim}}(t_j)^\eta\right) \epsilon_j, \tag{2.25}$$

where $\epsilon_j \overset{i.i.d.}{\sim} N(0,1)$. The spread of these noise models on PK and PD observations can be seen in Figure 2.10.

## 2.4 The population model

There are three approaches to inferring parameters of a system that describes multiple individuals. The first is the naive pooled (or fixed-effects) model. In this approach there is one set of parameters for the model which is assumed to describe every individual. Any deviation from the deterministic model with these parameters is incorporated into the noise and is independent of the individual. In reality, a biological response is likely to vary between individuals. A no-pooling (or two-stage) model tackles this by allowing each individual to have a different independent set of parameters to describe their response. This assumes that each individual's parameters can be determined by the observations of that individual and no information can be gained from the observations from other individuals. Parameters describing the whole population are then derived from these. However, in pharmaceutical settings it is common to only have a few observations per individual, meaning there is not sufficient information from a single individual to determine the parameters of the system. Also, the individual-level parameters are likely to not be fully independent of each other and useful information can be gained from the the whole population. These approaches can be combined into a mixed-effects (also known as partial-pooled

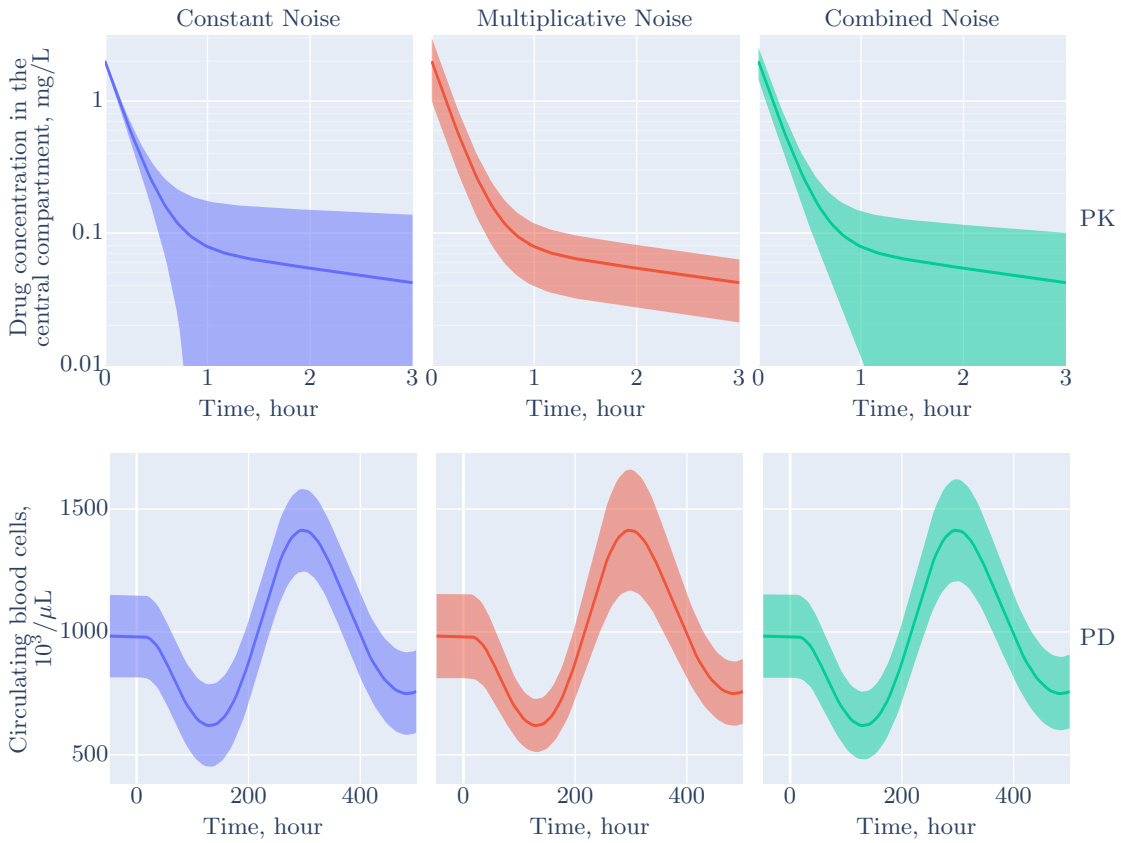



Figure 2.10: The simulation outputs of the two-compartment linear PK model (top) from Equations 2.2-2.3, and the Friberg PD model (bottom), from Equation 2.17-2.21. The parameters for these models are the same ones used in Figures 2.2 and 2.9. The shaded region signifies the area within one standard deviation of the simulated result when using constant Gaussian noise, Equation 2.23 (left) with $\sigma_{PK,c} = 0.096$, $\sigma_{PD,c} = 167.7$; multiplicative Gaussian noise, Equation 2.24 (middle) with $\sigma_{PK,m} = 0.5$, $\sigma_{PD,m} = 0.17$, $\eta = 1$; or combined Gaussian noise, Equation 2.25 (right) with $\sigma_{PK,c} = 0.048$, $\sigma_{PD,c} = 83.9$, $\sigma_{PK,m} = 0.25$, $\sigma_{PD,m} = 0.087$, $\eta = 1$.

or hierarchical) model. These models assume that each individual has an individual set of parameters determined from a population distribution of parameters. This population distribution is further described by hyper-parameters. When parameter inference is performed on a mixed-effects model, the hyper-parameters are inferred alongside the individual-level parameters. [16]

To extend a deterministic model with parameters $\theta \in \Theta$ into a mixed-effects model, the parameters to vary, $\psi \subseteq \theta$, need to be determined. It may not be appropriate to allow all parameters to vary as this introduces many additional parameters and can lead to over-fitting. We then let $\theta_i = (\psi_i, \theta_{-\psi}) \in \Theta$ be the parameter set for individual $i \in [1, n_{\text{ind}}]$. Here, $\theta_{-\psi}$ are the fixed-effects parameters, and $\psi_i = f(\psi_{typ}, \eta_i)$ are the mixed-effects parameters with fixed effect (or typical value) $\psi_{typ}$ and random effect $\eta_i$. The random effect is determined by a probability distribution, $\eta_i \sim Z(\Omega)$, for inter-individual variability parameters $\Omega$. The hyper parameters of this model are $\{\psi_{typ}, \theta_{-\psi}, \Omega\}$, in addition to the individual-level parameters. Most papers which have mixed-effects models start with the a complicated population model where $\psi$ encompasses most or all of $\theta$. They then drop one parameter, $\theta_k$ at a time from $\psi$. If the fit with $\theta_k$ as a fixed-effects parameter is not significantly worse, then $\psi$ is updated with $\theta_k$ excluded [7, 17]. Alternatively, analysis could start with the simple case, $\psi$ is a single parameter, and slowly increase complexity by adding one parameter at a time to $\psi$ until no parameters can significantly improve the fit.

Commonly in PKPD models, a log-normal distribution is used for the distribution of individual-level parameters, i.e.

$$\psi_i = \psi_{\text{typ}} \cdot \exp(\eta_i),$$

where $\psi_{\text{typ}}$ is the population typical value, $\psi_i$ is the individual parameter value, and $\eta_i \sim N(0, \Omega)$ are the random effects. This model ensures the individual parameter values stays above zero as is required for all parameters of the models outlined in this thesis. For simplicity, the covariance matrix $\Omega$ is often restricted to a diagonal matrix with values $\Omega_{k,k} = \omega^2_{\psi_k}$, the inter-individual variance of parameter $\psi_k$. [7, 33, 17]

## 2.5 Conclusion

In order to build PKPD models, information about the system is needed which depends on the structure of the data. For the PK model, a complex model such as the physiologically-based or general $n$-compartment PK model is simplified down to the mechanism could be supported by the data. For our case this is the two-compartment

PK model, as will be seen in section 4. For the PD part of the model, the opposite approach has been taken, the simple indirect response model has been further built upon to include additional dynamics commonly observed in the data. For this case study, this has lead to the the Friberg PD model. In many PKPD modelling studies, the noise model and the population model are not explored as thoroughly but are equally as important to the overall model. Thus, in this thesis I concentrate on comparing these aspects of the model while building the methodology of model analysis. I will be using both the fixed-effects population model, as a simple model to test our methodology on, as well as the mixed-effects population model, which is the most commonly used population model in PKPD modelling. I will analyse all 3 noise models for the fixed-effects population model. This analysis will then inform my choice of noise model for the mixed-effects model analysis. I will start the mixed-effects modelling with $\psi = [V_c]$ and iteratively add a single parameter at a time. This will hopefully prioritise simplicity in the model and be easier to analyse. I will use a log-normal distribution for $\psi_i$ with diagonal covariance matrix.

# Chapter 3

# Parameter identifiablity

As seen in Chapter 2, the statement of the forward problem comes naturally: given a model and a set of parameters, $\theta = (K_{cl}, V_c, K_{\text{tr}}, s, \ldots)$, what is the resulting time-profile of the variable, $R$. However, in our case, we have the inverse problem: given a data-set, which model and set of parameters can best describe the observations. Our first step in solving the inverse problem is to ensure it can be solved, i.e. given a model and data, can the parameters of the model be determined. This is known as the identifiabilty of the model, and there are two different non-identifiabilities that might be present which can affect the results of parameter inference. The first is structural non-identifiablity, where the model is defined in such a way that different parameters can produce the same output. The other is practical non-identifiability, where different parameter values cannot be discerned in parameter inference due to the sparsity and noisiness of data. Practical identifiability will be analysed in Chapter 6. In this chapter, before any data is introduced, the structural identifiability of the model can be determined. I will formally define structural identifiability, discuss methods of determining the identifiability for both fixed-effects and mixed-effects models, and run identifiability analysis on the fixed-effects model. In the fixed effects case, we are concerned with whether the deterministic parameters are identifiable when given infinite noiseless data, i.e. a continuous output curve. However mixed-effects models are more complicated, as for these models we are concerned with whether the population parameters are identifiable when given infinite patients with infinite noiseless data, i.e. a probability distribution of the output that changes over time. As this population based modelling is not as common as systems of ODEs, there is no software available to perform structural identifiability analysis. Thus I could not perform the analysis for the full PKPD mixed-effects model.

## 3.1 Structural identifiability

The structural identifiabilty problem assumes that there is infinite noiseless data, and the output of the system is fully known. A fixed-effects model is structurally locally identifiable if for all $\theta \in \Theta$, there exists a neighbourhood $N_\theta \subset \Theta$ of $\theta$ such that, if $\bar{\theta} \in N$ satisfies

$$y(\theta, t) = y\left(\bar{\theta}, t\right) \ \forall t,$$

where $y(\theta, t)$ then

$$\theta = \bar{\theta}.$$

If additionally $N_\theta = \Theta \ \forall \theta$ then the model is globally identifiable. Otherwise, the the model is unidentifiable [29]. One method to determine the structural identifiability is performed by taking the Taylor expansion of the output (if it exists) around $t = 0$ [31],

$$y(\theta, t) = y(\theta, t) + y^{(1)}(\theta, 0)\, t + y^{(2)}(\theta, 0)\, \frac{t^2}{2!} + \dots.$$

The equality $y(\theta, t) = y\left(\bar{\theta}, t\right)$ is then equivalent to $y^{(k)}(\theta, 0) = y^{(k)}\left(\bar{\theta}, 0\right)$ for $k = 0, 1, \dots$. For a linear system of ODEs, such as the linear PK models defined in section 2.1, it is sufficient to check the equations up to $k = 2n_{\text{states}} - 1$, where $n_{\text{states}}$ is the number of states in the system of ODEs.

Taking the two-compartment PK model (Equations 2.2-2.3) with parameters $\theta = (V_c, K_{\text{cl}}, V_1, K_1)$, the output is $y(\theta, t) = \frac{A_c(t)}{V_c}$ and the initial values are $A_c(0) = d$, the known dose amount, and $A_1(0) = 0$.

Now, let $\theta, \bar{\theta} \in \Theta$ be such that $y(\theta, t) = y\left(\bar{\theta}, t\right) \ \forall t \geq 0$. Then the first term of the Taylor's series gives

$$\begin{aligned}
& y(\theta, 0) = y\left(\bar{\theta}, 0\right) \\
\implies & \frac{A_c(0)}{V_c} = \frac{A_c(0)}{\bar{V}_c} \\
\implies & \frac{d}{V_c} = \frac{d}{\bar{V}_c} \\
\implies & V_c = \bar{V}_c \qquad \text{if } d \neq 0.
\end{aligned}$$

Looking at the second term of the series,

$$\begin{aligned}
& y^{(1)}(\theta,0) = y^{(1)}\left(\bar{\theta},0\right) \\
\implies & \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\left(\frac{A_c(t)}{V_c}\right) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\left(\frac{A_c(t)}{\bar{V}_c}\right) \\
\implies & \frac{1}{V_c}\left(-K_{\mathrm{cl}}\frac{A_c(0)}{V_c} - K_1\left(\frac{A_c(0)}{V_c} - \frac{A_1(0)}{V_1}\right)\right) \\
& \qquad = \frac{1}{\bar{V}_c}\left(-\bar{K}_{\mathrm{cl}}\frac{A_c(0)}{\bar{V}_c} - \bar{K}_1\left(\frac{A_c(0)}{\bar{V}_c} - \frac{A_1(0)}{\bar{V}_1}\right)\right) \\
\implies & -\frac{d}{V_c^2}\left(K_{\mathrm{cl}} + K_1\right) = -\frac{d}{\bar{V}_c^2}\left(\bar{K}_{\mathrm{cl}} + \bar{K}_1\right) \\
\implies & K_{\mathrm{cl}} + K_1 = \bar{K}_{\mathrm{cl}} + \bar{K}_1, \qquad \text{as } V_c = \bar{V}_c.
\end{aligned}$$

The final two terms we need to examine (as it is sufficient to check up to $k = 2n_{\mathrm{states}} - 1 = 3$) are

$$\begin{aligned}
& y^{(2)}(\theta,0) = y^{(2)}\left(\bar{\theta},0\right) \\
\implies & \frac{d\left(K_c + K_1\right)^2}{V_c^3} + \frac{dK_1^2}{V_c^2 V_1} = \frac{d\left(\bar{K}_c + \bar{K}_1\right)^2}{\bar{V}_c^3} + \frac{d\bar{K}_1^2}{\bar{V}_c^2\bar{V}_1} \\
\implies & \frac{K_1^2}{V_1} = \frac{\bar{K}_1^2}{\bar{V}_1}, \qquad \text{as } V_c = \bar{V}_c \text{ and } K_{\mathrm{cl}} + K_1 = \bar{K}_{\mathrm{cl}} + \bar{K}_1,
\end{aligned}$$

and

$$\begin{aligned}
& y^{(3)}(\theta,0) = y^{(3)}\left(\bar{\theta},0\right) \\
\implies & \frac{d\left(K_c + K_1\right)^3}{V_c^4} + \frac{2d\left(K_c + K_1\right)K_1^2}{V_c^3 V_1} + \frac{dK_1^3}{V_c^2 V_1^2} \\
& \qquad = \frac{d\left(\bar{K}_c + \bar{K}_1\right)^3}{\bar{V}_c^4} + \frac{2d\left(\bar{K}_c + \bar{K}_1\right)\bar{K}_1^2}{\bar{V}_c^3\bar{V}_1} + \frac{d\bar{K}_1^3}{\bar{V}_c^2\bar{V}_1^2} \\
\implies & \frac{K_1}{V_1} = \frac{\bar{K}_1}{\bar{V}_1}, \qquad \text{as } V_c = \bar{V}_c,\ K_{\mathrm{cl}} + K_1 = \bar{K}_{\mathrm{cl}} + \bar{K}_1 \text{ and } \frac{K_1^2}{V_1} = \frac{\bar{K}_1^2}{\bar{V}_1}.
\end{aligned}$$

$\theta = \bar{\theta}$ is the only solution to the above equations, which means this model is globally structurally identifiable.

For the Friberg PD model, I have utilised STRIKEGOLDD, a toolbox for MATLAB that has implemented a variety of methods that determines structural identifiability. This is more efficient than doing the above steps by hand or using the symbolic solver in MATLAB, especially for larger and more complicated systems of ODEs. This software has determined that the Friberg model is identifiable.

## 3.2 Structural identifiability of Mixed-Effects Models

In the mixed-effects model, structural identifiability is determined when infinite individuals are sampled from the population distribution, and infinite observations taken from each individual. The output in this case is a probability distribution that changes over time. Thus to formally define structural identifiability of the model, we let the probability distributions $Y\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, t\right)$ and $Y\left(\{\bar{\psi}_{typ}, \bar{\theta}_{-\psi}, \bar{\Omega}\}, t\right)$ be the same, i.e.

$$\mathrm{E}\left[Y\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)^j\right] = \mathrm{E}\left[Y\left(\{\bar{\psi}_{typ}, \bar{\theta}_{-\psi}, \bar{\Omega}\}, 0\right)^j\right], \forall j \in \mathbb{N}$$

and

$$\mathrm{Cov}\left(Y\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)\right) = \mathrm{Cov}\left(Y\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)\right).$$

If this implies

$$\{\psi_{typ}, \theta_{-\psi}, \Omega\} = \{\bar{\psi}_{typ}, \bar{\theta}_{-\psi}, \bar{\Omega}\}$$

for $\{\bar{\psi}_{typ}, \bar{\theta}_{-\psi}, \bar{\Omega}\} \in N$, a neighbourhood $N \subset \Theta$ of $\{\psi_{typ}, \theta_{-\psi}, \Omega\}$, then the model is locally identifiable. If additionally $N = \Theta$ then the model is globally identifiable. Otherwise, the the model is unidentifiable [29].

The taylor expansion can be used again in this case [31], thus the above equality is true if and only if

$$\mathrm{E}\left[Y^{(k)}\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)^j\right] = \mathrm{E}\left[Y^{(k)}\left(\{\bar{\psi}_{typ}, \bar{\theta}_{-\psi}, \bar{\Omega}\}, 0\right)^j\right]$$

and

$$\mathrm{Cov}\left(Y^{(k)}\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)\right) = \mathrm{Cov}\left(Y^{(k)}\left(\{\psi_{typ}, \theta_{-\psi}, \Omega\}, 0\right)\right), \forall j \in \mathbb{N}, k \in \mathbb{N}^0.$$

However, the current software for determining structural identifiability is designed for systems of ODEs, not for probability distributions. Attempting by hand takes a long time and thus these computational methods need to be updated for mixed-effects models. For this thesis, I will observe that if a fixed-effects model is identifiable then the individual-level parameters, $\theta_i$, of the mixed-effects model is identifiable. As there are infinite individuals, the full distribution of each parameter is known and thus if the distribution is well defined, the population is identifiable.

## 3.3 Conclusion

In conclusion we can confirm that the mechanistic PKPD models are globally structurally identifiable. However, these models are commonly used in the literature and

thus would be expected to be identifiable. In this thesis, I am specifically exploring the noise models and population models. The noise model cannot be determined to be structurally identifiable given that structural identifiability requires noiseless data. With the population model, I have found that it can be difficult to analyse a mixed-effects model when there is a lot of complexity in the mechanistic model. However, with a well-defined mixed-effects model and an identifiable mechanistic model, the mixed-effects model would also be identifiable. Additionally, as will be seen in Chapter 6, methods that assess practical identifiability can also identify local structural identifiability.

# Chapter 4

# Data

As discussed in Chapter 2, many decisions made in mathematical modelling are a result of observations made, in addition to the purpose of the model. As will be seen in subsequent sections, observations are required to analyse and refine the models, in order to produce reliable predictions. In this thesis, I will be using both simulated data, to test and refine the methods used throughout, and experimental data provided by my colleague at Roche. In this section I will describe these datasets and explain how the experimental data has lead to the choices made in the deterministic model.

## 4.1 Experimental data

The experimental data used in this report is from a preclinical study on rats. The single-dose study involves dosing rats with five different anti-tumour drugs at differing dose levels. Observations on drug concentrations in the blood stream and nine different blood cell counts are provided in the data. In this report I consider the drug Docetaxel and the concentration of platelets in the blood. The raw PK and platelet data gathered for Docetaxel can be seen in Figure 4.1.

This data informs the choice of model to use to simulate these results. I have chosen a two-compartment linear PK model,

$$\frac{\mathrm{d}A_c}{\mathrm{d}t} = -K_{\mathrm{cl}}\frac{A_c}{V_c} - K_1\left(\frac{A_c}{V_c} - \frac{A_1}{V_1}\right), \tag{4.1}$$

$$\frac{\mathrm{d}A_1}{\mathrm{d}t} = K_1\left(\frac{A_c}{V_c} - \frac{A_1}{V_1}\right), \tag{4.2}$$

as Figure 2.3 shows clearly at least two linear sections and the dose-normalised curves

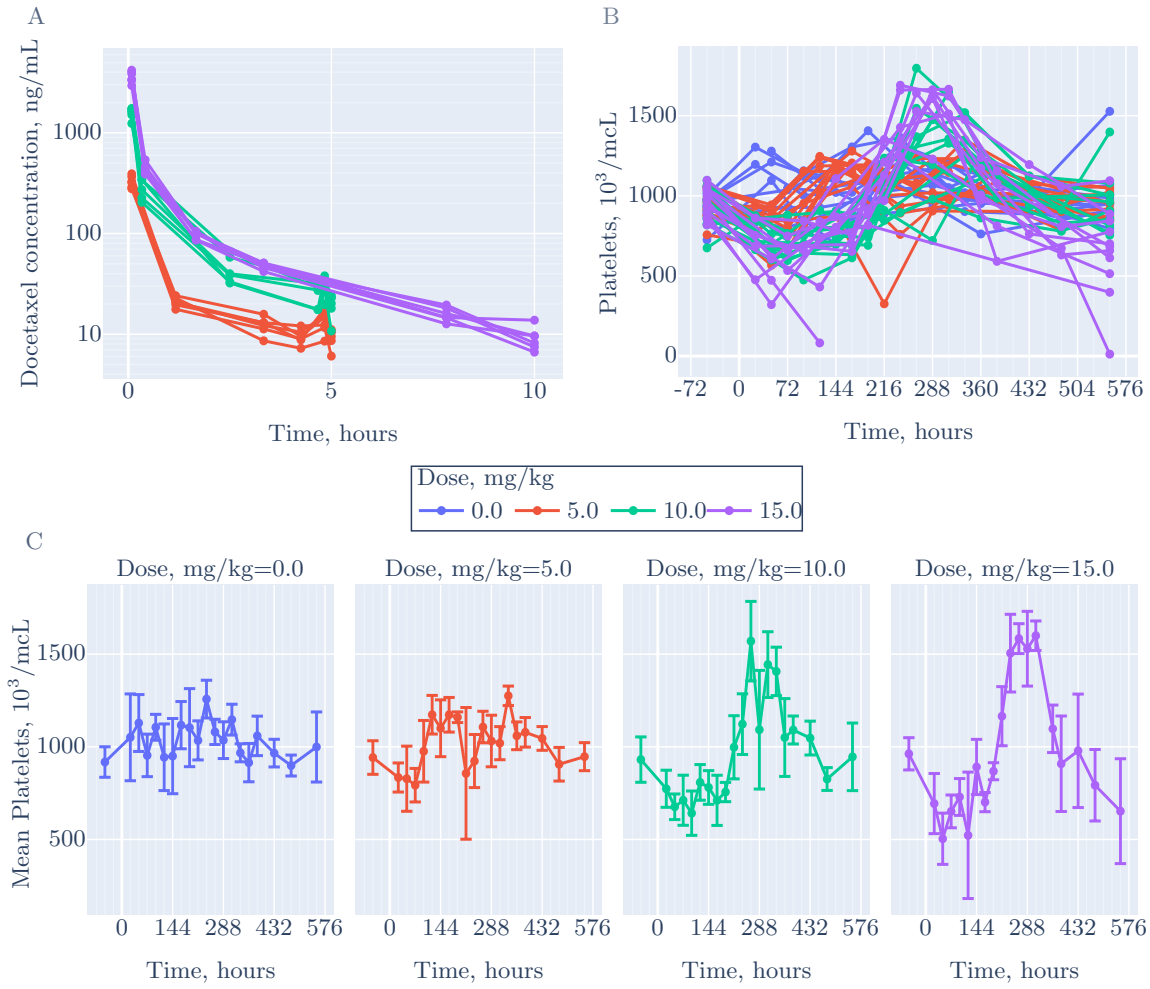



Figure 4.1: Overview of data for Docetaxel. A) PK data of the concentration of Docetaxel in the blood over time, plotted on a log scale. B) PD data of the time course of platelet concentration in each individual rat. C) PD data of the platelet concentration averaged over individuals in each dose group. Docetaxel or saline solution was administered at time t=0.

overlap for the majority of points. I have also selected the Friberg PD model,

$$\frac{\mathrm{d}P}{\mathrm{d}t} = K_{\mathrm{prol}}\left(1 - E_{\mathrm{drug}}\left(t\right)\right)\left(\frac{R_0}{R}\right)^{\gamma} P - K_{\mathrm{tr}}P, \tag{4.3}$$

$$\frac{\mathrm{d}T_1}{\mathrm{d}t} = K_{\mathrm{tr}}P - K_{\mathrm{tr}}T_1, \tag{4.4}$$

$$\frac{\mathrm{d}T_2}{\mathrm{d}t} = K_{\mathrm{tr}}T_1 - K_{\mathrm{tr}}T_2, \tag{4.5}$$

$$\frac{\mathrm{d}T_3}{\mathrm{d}t} = K_{\mathrm{tr}}T_2 - K_{\mathrm{tr}}T_3, \tag{4.6}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = K_{\mathrm{tr}}T_3 - K_{\mathrm{out}}R, \tag{4.7}$$

as it is the model that can exhibit both the rebound effect and the delayed drug response that can be seen in this data. I have used a linear drug effect model, $E_{\mathrm{drug}} = SC_0\left(t\right)$ to keep the model relatively simple. I assume the PK data can be captured with a multiplicative noise model, as is commonly used in PK modelling, and I will compare the different noise models for the PD data when analysing the fixed-effects population model. The best noise model from this analysis will then be used in the mixed-effects population model.

The data provided has time courses for each individual and thus some normalisation could occur such as dividing through by the initial measurement. This may improve the accuracy of determining the drug effect and feedback, as the differences between any two points will not be affected by the different starting points. However, elimination of blood cells relies on the concentration rather than the ratio with the base level. This means, for the fixed effects model, simulations will need to be performed for each individual rather than for each dose and this increases the time taken to analyse. Additionally, there is some error in the first observation which is not known and thus it cannot be assumed that it is equivalent to $R_0$. This may complicate the analysis performed. The range in initial points can be captured by a mixed-effects model for $R_0$ and I explore this in Chapter 7. Thus, in keeping with the approach that Friberg *et. al.* have taken, I have decided not to alter the data in this manner.

## 4.2 Simulated data

I will also test and explore the methods of analysis used in this report on simulated data sets. These data-sets will be generated from one of the following:

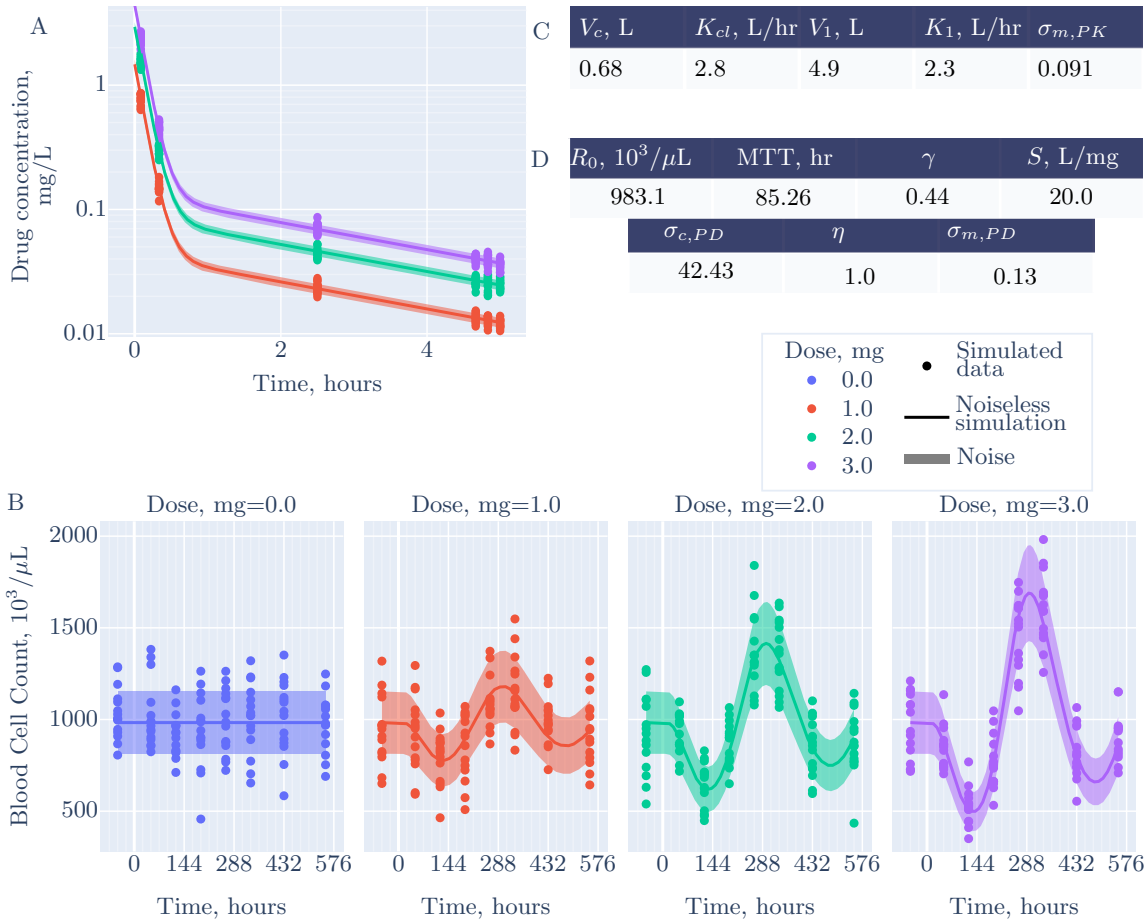


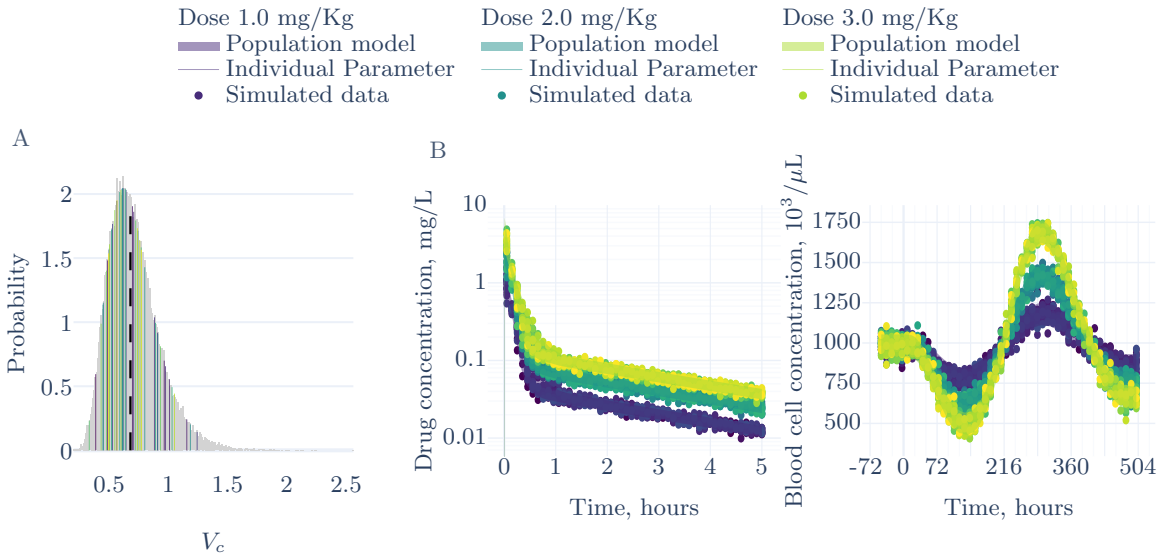

Figure 4.2: Overview of the simulated data. A) Plot of PK data, the drug concentration in the central compartment over time, plotted on a log scale. B) Plot of PD data, the time course of blood cell concentration. Tables of the C) PK and D) PD parameters used in simulating the data. Drug or placebo was administered at time $t = 0$. The shaded region signifies the area within one standard deviation of the simulated result.

1. Fixed-effects two-compartment linear PK model (Equations 2.2-2.3) with multiplicative noise, and the Friberg PD model (Equations 2.17-2.21) with combined multiplicative and constant noise (Equation 2.25). The observed variables in the data are the concentration of drug in the central compartment and the concentration of circulating blood cells. One data-set is generated with 15 individuals per dose, under 3 doses and a control, at time points that match those in the experimental data. Figure 4.2 shows this data and the parameters used to simulate it.

2. Mixed-effects Two-compartment linear PK model (Equations 2.2-2.3) with multiplicative noise, and the Friberg PD model (Equations 2.17-2.21) with constant noise (Equation 2.23). The mixed-effects parameters are $\psi = [V_c]$. The observed variables in the data are the concentration of drug in the central compartment and the concentration of circulating blood cells. One data-set is generated with 15 individuals per dose, under 3 doses, with 50 observations per individual. Figure 4.3 shows this data and the parameters used to simulate it.

3. Mixed-effects one-compartment linear PK model (Equation 2.1) with multiplicative noise. The observed variable in the data is the concentration of drug in the central compartment. Eight data-sets are generated. The mixed-effects parameters are $\psi = [V_c]$ for the first four data-sets and $\psi = [V_c, K_{cl}]$ for the last four data-sets. Figure 4.4 shows two of these data-sets and the parameters used to simulate it.

The noise model for the data-set in category 1 was chosen as it is initially assumed by Friberg *et. al.* when performing the analysis. However, the analysis in Chapter 6 shows practical identifiability problems and thus, for the data-set in category 2, was reduced to a constant noise model. The parameters used to produce the simulated data in category 1 and 2 are similar to those from the parameter inference of the experimental data (see Chapter 5) to ensure it has similar properties to the data. The third category initially had $V_c$ and $K_{cl}$ equal to the same values as in category two. However, this created dynamics that were difficult to solve numerically, as the drug concentration decreased incredibly quickly into values that would not be experimentally detectable. This difference in scales between the two models is likely due to the difference in total volumes that the drug is distributed between, being more than halved with the removal of the peripheral compartment. To combat this I instead set $V_c$ in the one-compartment model to the sum of $V_c$ and $V_1$ from the two-compartment model.




C

| $K_{cl}$, L/hr | $V_{c,typ}$, L | $\omega_{V_c}$ | $K_1$, L/hr | $V_1$, L | $\sigma_{m,PK}$ |
|---|---|---|---|---|---|
| 2.8 | 0.68 | 0.3 | 2.3 | 4.9 | 0.091 |

D

| $S$, L/mg | $R_0$, $10^3/\mu$L | $\gamma$ | MTT, hr | $\sigma_{c,PD}$ |
|---|---|---|---|---|
| 20.0 | 983.1 | 0.44 | 85.26 | 0.13 |

Figure 4.3: A) Graph showing the probability distribution of individual parameter, $V_{c,i}$, of the population. This is the Log-normal distribution with $V_{c,typ} = 0.68$, shown by the black dashed line and $\omega_{V_c} = 0.3$. The coloured lines show the values for each individual in our sample used to generate the data. B) The simulated data generated from these values. The shaded region signifies the 5-th to 95-th percentile expected from the population model. Tables of the C) PK and D) PD parameters used in simulating the data. Drug or placebo was administered at time $t = 0$.

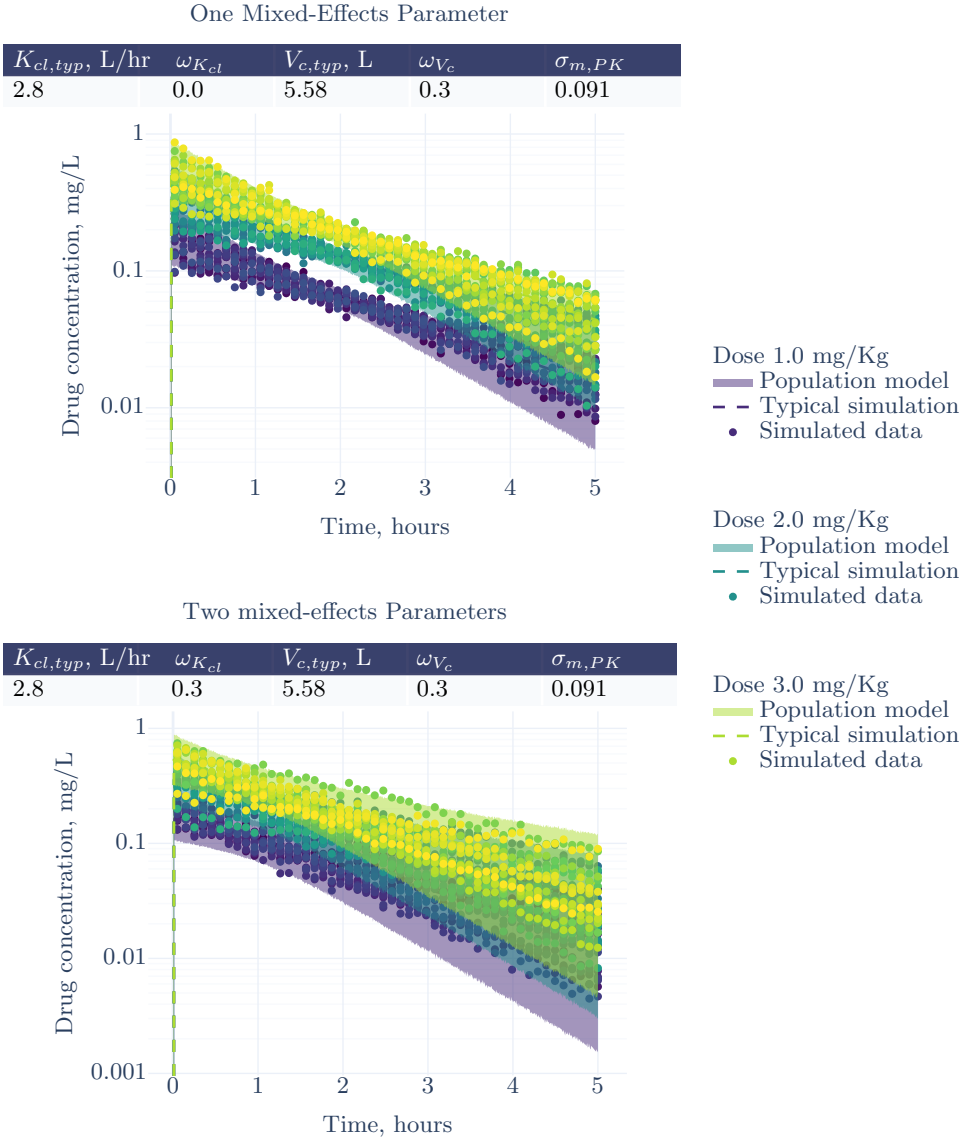



Figure 4.4: Simulated PK data-sets 1 and 5 (one and two mixed-effects parameters respectively). Table of the parameters used to generate the data and graph of the drug concentration over time. Shaded region shows the 5-th to 95-th percentile expected from the population model. Y-axis is on a log scale.

The data within category 3 (Figure 4.4), shows the differences in the spread of the data caused by the introduction of different parameters as mixed-effects, on a log scale. The spread is increased at the start and the end of the simulation when $V_c$ has variability. Introducing variability to the parameter $K_{cl}$ further increases the spread, with increasing effect at later times.

## 4.3 Conclusion

I have generated three different categories of data-sets, with which I will be testing my methods and analysing, alongside the experimental data. This simulated data is generated such that it is similar to the experimental data. However, with the last two data categories, I increased the number of data-points, so that when testing the mixed-effects analysis, results could be more reliable. The last data category was also generated as the simplest PK model, to test the efficiency and reliability of methods with fewer parameters. This ensures that multiple runs of the same method do not take an incredibly long time to compute.

# Chapter 5

# Maximum likelihood estimation

Given appropriate data and identifiable models, the inverse problem, as described in Chapter 1, can be solved by parameter inference and model selection. Then, after the inverse problem has been solved, I can explore the forward problem to make predictions on the effects of differing doses in Chapter 9. In this chapter I will be using maximum likelihood estimation and maximum-a-posteriori estimation to perform parameter inference. I will explain these methods and use them to perform parameter inference on the data-sets described in Chapter 4. I will also compare these to Monolix, a software used for parameter inference in mixed-effects PKPD models.

A common approach to parameter inference is to maximise the probability that the data, $X^{\text{obs}}$, was generated by the model with parameters, $\theta \in \Theta$, over the parameter space $\Theta$. This likelihood, $P\left(X^{\text{obs}}|\theta\right)$ can be calculated as the product of point-wise likelihoods,

$$P\left(X^{\text{obs}}|\theta\right) = \prod_{x_j^{\text{obs}} \in X^{\text{obs}}} P\left(x_j^{\text{obs}}|\theta\right).$$

Alternatively, a computationally easier but equivalent optimisation is the log-likelihood,

$$\log\left(P\left(X^{\text{obs}}|\theta\right)\right) = \sum_{x_j^{\text{obs}} \in X^{\text{obs}}} \log\left(P\left(x_j^{\text{obs}}|\theta\right)\right).$$

As all of the noise models explored in this report are Gaussian, each of these observations are expected to be normally distributed around the simulation, i.e. $x_j^{\text{obs}} \sim N\left(x^{\text{sim}}\left(t_j\right), \sigma_j^2\right)$ where $\sigma_j$ is the standard deviation of the observation at time $t = t_j$. Thus the point-wise likelihood can be formulated as a Gaussian distribution,

$$\begin{aligned} P\left(x_j^{\text{obs}}|\theta\right) &= P\left(x_j^{\text{obs}}|x^{\text{sim}}\left(t_j\right) = x^{\theta}\left(t_j\right), \sigma_j\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\left(x_j^{\text{obs}} - x^{\theta}(t_j)\right)^2 \sigma_j^{-2}}, \end{aligned}$$

where $x^\theta$ is the simulated dynamical model under parameters $\theta$. The point-wise log-likelihood would then be

$$\log\left(P\left(x_j^{\text{obs}}|\theta\right)\right) = -\frac{1}{2}\log\left(2\pi\right) - \log\left(\sigma_j\right) - \frac{1}{2}\left(\frac{x_j^{\text{obs}} - x^\theta\left(t_j\right)}{\sigma_j}\right)^2,$$

However, $\sigma_j$ differs depending on the noise model. If a constant noise is assumed, $\sigma_j = \sigma_c$ and the log-likelihood is

$$\log\left(P\left(x_j^{\text{obs}}|\theta\right)\right) = \sum_{x_j^{\text{obs}} \in X^{\text{obs}}} \left(-\frac{1}{2}\log\left(2\pi\right) - \log\left(\sigma_c\right) - \frac{1}{2}\left(\frac{x_j^{\text{obs}} - x^\theta\left(t_j\right)}{\sigma_c}\right)^2\right).$$

For multiplicative noise, $\sigma_j = \sigma_m x^{\text{sim}}\left(t_j\right)^\eta$ and the log-likelihood is

$$\log\left(P\left(x_j^{\text{obs}}|\theta\right)\right) = \sum_{x_j^{\text{obs}} \in X^{\text{obs}}} \left(-\frac{1}{2}\log\left(2\pi\right) - \log\left(\sigma_m x^{\text{sim}}\left(t_j\right)^\eta\right) - \frac{1}{2}\left(\frac{x_j^{\text{obs}} - x^\theta\left(t_j\right)}{\sigma_m x^{\text{sim}}\left(t_j\right)^\eta}\right)^2\right).$$

And in the case of combined multiplicative and constant noise, $\sigma_j = \sigma_c + \sigma_m x^\theta\left(t_j\right)^\eta$ and the log-likelihood is

$$\log\left(P\left(x_j^{\text{obs}}|\theta\right)\right) = \sum_{x_j^{\text{obs}} \in X^{\text{obs}}} \left(-\frac{1}{2}\log\left(2\pi\right) - \log\left(\sigma_c + \sigma_m x^{\text{sim}}\left(t_j\right)^\eta\right) - \frac{1}{2}\left(\frac{x_j^{\text{obs}} - x^\theta\left(t_j\right)}{\sigma_c + \sigma_m x^\theta\left(t_j\right)^\eta}\right)^2\right).$$

Maximising the likelihood with the constant Gaussian noise model, is equivalent to minimising the sum of squares error,

$$f\left(\theta\right) = \sum_j \left(x^\theta\left(t_j\right) - x_j^{\text{obs}}\right)^2,$$

as

$$\log\left(P\left(X^{\text{obs}}|\theta\right)\right) = A - Bf\left(\theta\right),$$

for constants $A = n_{\text{obs}}\left(-0.5\log 2\pi - \log\sigma_c\right)$ and $B = 0.5\sigma_c^{-2}$, where $n_{\text{obs}}$ is the number of observations. However, $f\left(\theta\right)$ is only concerned with the deterministic model parameters and minimising this function does not provide $\sigma_c$.

To find the optimum of these likelihoods I used the CMA-ES (Covariance Matrix Adaptation Evolution Strategy) algorithm to search the parameter space, implemented by the PINTS package for Python [8]. This searching algorithm works by having a population of points, $Y^{(g)} \subset \Theta$, and iteratively generating new populations

based on the output of the objective function, the function to optimise, at the points in the previous population. This new population, $Y^{(g+1)}$, is selected using a normal distribution based on the covariance matrix:

$$\theta_k^{(g+1)} = \mathbf{m}^{(g)} + h^{(g)}\nu_k, \qquad \forall\theta_k^{(g+1)} \in Y^{(g+1)},$$

where $\nu_k \sim \mathcal{N}\left(\mathbf{0},\, \mathbf{C}^{(g)}\right)$ and $\mathbf{m}^{(g)} = \sum_{\theta_x \in B^{(g)}} w_x\theta_x$, a weighted average over the subset $B^{(g)} \subset Y^{(g)}$ of the best points in the sample according to our objective function. The weights are such that the points in the population are weighted more highly; $h^{(g)}$ is the step size that is adapted at each iteration; and $\mathbf{C}^{(g)}$ is an estimation of the covariance matrix using the $g$th and $(g-1)$th population [22].

There are 12 parameters to estimate in the fixed-effects Friberg PKPD model: $K_{\text{cl}}$, $K_1$, $V_c$, $V_1$, $\sigma_{\text{m}}$, $R_0$, MTT, $\gamma$, $s$, $\sigma_{\text{c,PD}}$, $\sigma_{\text{m,PD}}$ and $\eta_{\text{PD}}$. This means there is a 12-dimensional parameter space to search. However to speed up the searching, I have split the parameter inference into two problems, the PK problem and the PD problem. The PK problem optimises the likelihood $P\left(X_{\text{PK}}^{\text{obs}}|\theta_{\text{PK}}, \sigma_c\right)$, over the five-dimensional parameter space $\Theta_{\text{PK}}$, with $\theta_{\text{PK}} = (K_c, K_1, V_c, V_1, \sigma_{\text{m}}) \in \Theta_{\text{PK}}$ and given the drug concentration over time data, $X_{\text{PK}}^{\text{obs}}$. The PD problem optimises the likelihood $P\left(X_{\text{PD}}^{\text{obs}}|\theta_{\text{PD}}, \sigma_c\right)$, over the seven-dimensional parameter space $\Theta_{\text{PD}}$, with $\theta_{\text{PD}} = (C_{\text{cell},}\, 0, \text{MTT}, \gamma, s, \sigma_{\text{c,PD}}, \sigma_{\text{m,PD}}, \eta_{\text{PD}}) \in \Theta_{\text{PD}}$ and given the blood cell concentration over time data, $X_{\text{PD}}^{\text{obs}}$. The PK problem can be decoupled from the PD problem as the PK output, the drug concentration in the central compartment, $C_c(t)$, is not affected by the PD output, the circulating blood cells, $R(t)$, or the PD parameters, $\theta_{\text{PD}}$. However, The PD output is affected by the PK output and parameters, $\theta_{\text{PK}}$, but if the PK problem is already solved, $\theta_{\text{PK}}$ can be fixed to that solution during the PD analysis. Thus the maximum likelihood estimate is found initially for the PK parameters and data. Then the PD parameter inference is done on the PD data using the drug concentration over time simulated from the PK model with optimum parameters.

Figure 5.1 shows the results of maximum likelihood estimation on category 1 simulated PD data. The choice of noise model seemed to have little impact on the inferred deterministic parameters, $(R_0, \text{MTT}, \gamma, s)$. However, the true noise parameters were not recovered in the inference. When combined noise was assumed, the maximum likelihood estimate for $\sigma_{c,\text{PD}}$ was 0.001, the lower bound for this parameter, and the maximum likelihood estimate for $\sigma_{m,\text{PD}}$ was 0.900403, close to the upper bound for this parameter. It seems, to compensate for this large value of $\sigma_{m,\text{PD}}$, $\eta$ was inferred much lower than its actual value.

A

| Parameter | Constant noise | Multiplicative noise | Combined noise | Actual parameters |
|---|---|---|---|---|
| $R_0$, $10^3/\mu$L | 987.92 | 988.26 | 988.26 | 983.1 |
| MTT, hr | 86.76 | 85.63 | 85.63 | 85.26 |
| $\gamma$ | 0.45 | 0.44 | 0.44 | 0.44 |
| $S$, L/mg | 19.32 | 19.25 | 19.25 | 20.0 |
| $\sigma_{c,PD}$ | 166.64 | . | 0.0 | 42.43 |
| $\eta$ | . | 0.76 | 0.76 | 1.0 |
| $\sigma_{m,PD}$ | . | 0.9 | 0.9 | 0.13 |
| likelihood | 2696.51 | 3330.7 | 3126.53 | . |

B

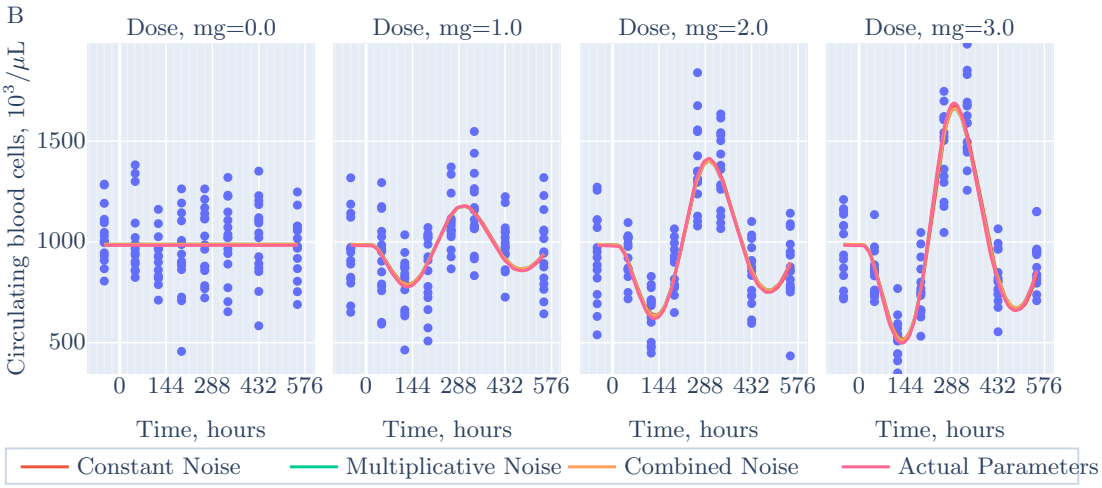
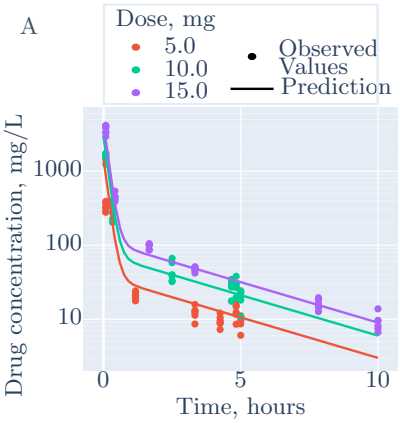



Figure 5.1: Results from the maximum likelihood estimation of the Friberg PD model (Equations 2.17-2.21) on simulated data. A) Table of PD parameters and the values acquired in optimisation when different noise models were assumed compared with the actual values used to create the simulated data. B) Comparison of the blood cell concentration observed in the simulated data, the actual curve used to create the data, and the curve produced using the parameters inferred when using different noise models.

I also inferred the parameters of the experimental data, assuming a fixed-effects two-compartment linear PK model (Equations 2.2-2.3) and the PD Friberg model (Equations 2.17-2.21) with linear drug effect. For the PK problem, a multiplicative noise was assumed, and for the PD problem constant, multiplicative and combined noise (Equations 2.23-2.25) were compared. The results can be see in Figure 5.2. All optimisations produced traces that fit well with data and found similar PD model parameters. Combined noise minimised the effect of the multiplicative part in the noise and inferred a similar value of $\sigma_{c,\mathrm{PD}}$ to the constant noise model, opposite of what happened in the simulated case. The multiplicative noise inferred the maximum value of $\sigma_{m,\mathrm{PD}}$, similar to the behaviour in the simulated case. This initially gives the impression that constant noise may be the most appropriate noise and that there may be identifiability issues in the multiplicative noise.

A




B

| $V_c$, L | $K_{cl}$, L/hr | $V_1$, L | $K_1$, L/hr | $\sigma_{m,PK}$ |
|---|---|---|---|---|
| 0.6838 | 2.7789 | 4.0892 | 1.7038 | 0.3127 |

C

| | Noise Model | | |
|---|---|---|---|
| Parameter | Constant | Multiplicative | Combined |
| $R_0$ | 989.96 | 989.7 | 989.58 |
| $MTT$ | 87.65 | 89.49 | 88.12 |
| $\gamma$ | 0.47 | 0.5 | 0.48 |
| $S$ | 15.03 | 12.85 | 14.77 |
| $\sigma_{c,\mathrm{PD}}$ | 193.77 | · | 193.08 |
| $\eta$ | · | 0.77 | 0.81 |
| $\sigma_{m,\mathrm{PD}}$ | · | 0.99 | 0 |
| likelihood | -2048.08 | -2381.33 | -2373.39 |

D

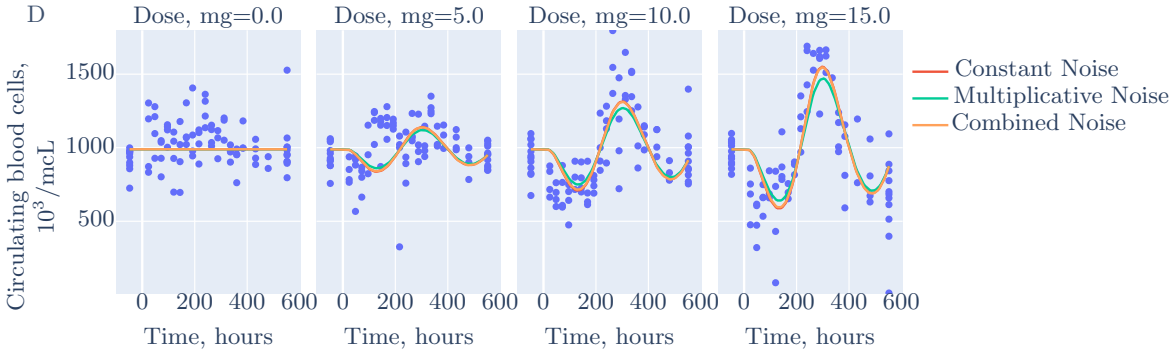



Figure 5.2: Results from the maximum likelihood estimation of the PKPD model parameters on experimental data. A) Table of PK parameters and the values acquired in optimisation. B) Prediction of Docetaxel concentration using the two-compartment linear PK model (Equations 2.2-2.3) and parameters shown in A) compared with the observed data. C) Table of PD parameters and the values acquired in optimisation when different noise models were assumed. D) Comparison of the platelet concentration observed in the experimental data, and the predictions produced using the Friberg model (Equations 2.17-2.21) and parameters shown in C).

## 5.1 Maximum-a-posteriori Estimation

An alternative parameter inference method is to find the maximum-a-posteriori estimate. This method maximises the probability distribution $P\left(\theta|X^{\text{obs}}\right)$, the probability that the parameter, $\theta$, could produce this data-set, for all parameters $\theta \in \Theta$. This probability distribution is defined as the posterior and is more directly the probability of interest to maximise, but is difficult to directly calculate. However, this posterior probability can be defined, using Bayes rule, in relation to the likelihood,

$$P\left(\theta|X^{\text{obs}}\right) = \frac{P\left(X^{\text{obs}}|\theta\right) P\left(\theta\right)}{P\left(X^{\text{obs}}\right)}, \tag{5.1}$$

where $P\left(X^{\text{obs}}|\theta\right)$ is the likelihood, as defined in Chapter 5; $P\left(\theta\right)$ is the prior, the probability that the parameters, $\theta$, can accurately represent this model before any data was collected; and $P\left(X^{\text{obs}}\right)$ is the marginal likelihood, the probability of the data, regardless of what parameters were used.

To maximise the posterior, assumptions about the form of the prior must be made. This can be a benefit to the modelling process as it allows the incorporation of information into the model outside of the data. For example if a parameter has particular bounds it must stay within, then this can be incorporated into the prior by setting $P\left(\theta\right) = 0$ outside these bounds. Other known information about a parameter, such as literature values, could also be incorporated by choosing a prior that has higher probability around these values. A relatively non-informative prior would be a uniform distribution where the boundaries are selected to ensure parameters stay within feasible values [48]. With non-informative priors, the maximum-likelihood and maximum-a-posteriori estimates should be the same, assuming the maximum-likelihood is within the set bounds.

Alternatively, information from a previous study on the same model, such as studies from previous phases of clinical trials for the same drug, can be used as a prior for the current study. This highly informative prior has the risk of dominating the posterior if there is similar or greater confidence in the parameters from the previous study than the current study. However confidence in parameters generally increases as size of the data set increases, and so larger studies have less risk of this. Alternatively, if using the model to predict the response in a single patient, it is useful to build a prior from a previous large clinical trial and for it to have an impact on the posterior, reducing the effect of noise in the single patient's observations.

The marginal likelihood, $P\left(X^{\text{obs}}\right)$, is not known but can be deduced from the fact that it is constant over $\theta$ and that the posterior is a probability distribution, thus

$$\int_{\Theta} P\left(\theta|X^{\text{obs}}\right) \, \mathrm{d}\theta = 1. \tag{5.2}$$

Continuing on from Bayes rule

$$\frac{\int_{\Theta} P\left(X^{\text{obs}}|\theta\right) P\left(\theta\right) \, \mathrm{d}\theta}{P\left(X^{\text{obs}}\right)} = 1,$$
$$\int_{\Theta} P\left(X^{\text{obs}}|\theta\right) P\left(\theta\right) \, \mathrm{d}\theta = P\left(X^{\text{obs}}\right).$$

However, when there are multiple parameters, the integral becomes difficult to calculate. On the other hand, as the marginal likelihood is constant, $P\left(\theta|X^{\text{obs}}\right) \propto P\left(X^{\text{obs}}|\theta\right) P\left(\theta\right)$ and the maximum of the posterior is the same as the maximum of $P\left(X^{\text{obs}}|\theta\right) P\left(\theta\right)$.

## 5.2 Transformation

To increase efficiency of the optimisation algorithms, and avoid complications with impossible parameter values being proposed and rejected, the parameter space can be transformed. This transformation, $q = T\left(\theta\right)$, will need to be such that $T^{-1}\left(\mathbf{q}\right) = \Theta$ for all $\mathbf{q} \in \mathbb{R}^{n_p}$, where $n_p$ is the number of parameters of the system. This can be easily done by setting

$$T\left(\theta\right) = \left(f_1\left(\theta_1\right), \ldots, f_n\left(\theta_{n_p}\right)\right),$$
$$f_k\left(\theta_k\right) = \begin{cases} \log\left(\theta_k - a_k\right) - \log\left(b_k - \theta_k\right), & \text{if } a_1 < \theta_k < b_1 \\ \log\left(\theta_k - a_k\right), & \text{if } a_k < \theta_k \\ \theta_k, & \text{if } \theta_k \in \mathbb{R} \end{cases}$$

where $\theta = \left(\theta_1, \ldots, \theta_{n_p}\right)$. In our model, all parameters are bounded below by zero thus the log transformation, $f_k\left(\theta_k\right) = \log\left(\theta_k\right)$, would be most appropriate.

To incorporate this into the maximum likelihood method, instead of maximising $LL\left(\theta\right)$ for $\theta \in \Theta$, $LL\left(T^{-1}\left(\mathbf{q}\right)\right)$ is maximised for $q \in \mathbb{R}^{n_p}$. Then when the optimum $q$ is returned, the MLE is $T^{-1}\left(\mathbf{q}\right)$.

## 5.3 Inference of Mixed-effects models

In a mixed-effects model the parameters are hierarchical which alters how the posterior is defined. In these types of models the posterior is

$$\begin{aligned} &P\left(\theta_1,\ldots,\theta_{n_i},\theta_{\text{typ}},\Omega|X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}\right) \\ &\quad\propto P\left(\theta_{\text{typ}},\Omega\right)\cdot P\left(\theta_1,\ldots,\theta_{n_i},|\theta_{\text{typ}},\Omega\right)\cdot P\left(X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}|\theta_1,\ldots,\theta_{n_i}\right), \end{aligned}$$

where

$$P\left(X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}|\theta_1,\ldots,\theta_{n_i}\right)=\prod_{i=1}^{n_i}P\left(X_i^{\text{obs}}|\theta_i\right)$$

is the likelihood,

$$P\left(\theta_1,\ldots,\theta_{n_i}|\theta_{\text{typ}},\Omega\right)=\prod_{i=1}^{n_i}P\left(\theta_i|\theta_{\text{typ}},\Omega\right)$$

is the prior, and $P\left(\theta_{\text{typ}},\Omega\right)$ is the hyperprior [28]. The log of this posterior,

$$\begin{aligned} &\log P\left(\theta_1,\ldots,\theta_{n_i},\theta_{\text{typ}},\Omega|X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}\right) \\ &\quad=\log P\left(\theta_{\text{typ}},\Omega\right)+\sum_{i=1}^{n_i}\left(LL_i\left(\theta_i\right)+\log P\left(\theta_i|\theta_{\text{typ}},\Omega\right)\right), \end{aligned}$$

is then optimised over for maximum-a-posteriori estimation, where $LL_i=\log P\left(X_i^{\text{obs}}|\theta_i\right)$ is the individual log-likelihood for individual $i$.

In our model $\psi$ is Log-normal with diagonal covariance, and thus the Log Prior for the individual-level parameters becomes

$$\begin{aligned} \log P\left(\psi_i,|\psi_{\text{typ}},\Omega\right)&=\sum_{\theta_k\in\psi}\log P\left(\theta_{k,i},|\theta_{\text{k, typ}},\omega_k^2\right) \\ &=\sum_{\theta_k\in\psi}\log\left(\frac{1}{\theta_{k,i}}\frac{1}{\sqrt{2\pi}\omega_k}\exp\left\{\left(-\frac{\left(\log\theta_{k,i}-\log\theta_{k,typ}\right)^2}{2\omega_k^2}\right)\right\}\right) \\ &=\sum_{\theta_k\in\psi}\left(-\log\theta_{k,i}-\log\left(\sqrt{2\pi}\omega_k\right)-\frac{\left(\log\theta_{k,i}-\log\theta_{k,typ}\right)^2}{2\omega_k^2}\right) \end{aligned}$$

I have also set the hyper prior for all parameters, $\{\psi_{\text{typ}},\theta_{-\psi},\omega\}$, as a log normal prior to ensure the lower bound of zero.

For the PKPD case with mixed-effects parameters $\psi$, the individual Log-likelihood is

$$LL_i\left(\theta_i\right)=LL_i\left(\psi_i,\theta_{-\psi}\right)=\sum_j PLL_{i,j}\left(\psi_i,\theta_{-\psi}\right)$$

The Friberg PKPD model with mixed-effects parameters $\psi$ would have two outputs. The first is the PK observation of the drug concentration in the central compartment for all individuals, $i$, $X_{\text{PK}}^{\text{obs}} = \{C_{c,i}(t_{j_1}) + \epsilon_{i,j_1} | t_{j_1} \in T_{PK}, 1 \leq i \leq n\}$, where $\epsilon_{i,j_1}$ is the observational error and, $T_{PK}$ is the set of times that the drug concentration is observed at. The second is the PD observation of circulating blood cells for all individuals, $i$, $X_{\text{PD}}^{\text{obs}} = \{R_i(t_{j_2}) + \epsilon_{i,j_2} | t_{j_2} \in T_{PD}, 1 \leq i \leq n\}$, where $\epsilon_{i,j_2}$ is the observational error and, $T_{PD}$ is the set of times that the blood cell count is observed at. In the PD mixed-effects model I am assuming a constant Gaussian noise model, and so the individual log-likelihoods are defined by

$$\begin{aligned}
\log\left(P\left(X_{\text{i}}^{\text{obs}} | \theta_{\text{i}}\right)\right) &= \sum_{x_{i,j_1}^{\text{obs}} \in X_{i,PK}^{\text{obs}}} \log\left(P\left(x_{i,j_1}^{\text{obs}} | \theta_i\right)\right) + \sum_{x_{i,j_2}^{\text{obs}} \in X_{i,PD}^{\text{obs}}} \log\left(P\left(x_{i,j_2}^{\text{obs}} | \theta_i\right)\right) \\
&= \sum_{x_{i,j_1}^{\text{obs}} \in X_{i,PK}^{\text{obs}}} \left( -\frac{\log(2\pi)}{2} - \log\left(\sigma_{PK} C_{c,i}(t_{j_1}, \theta_i)\right) - \frac{1}{2}\left(\frac{x_{i,j_1}^{\text{obs}} - C_{c,i}(t_{j_1}, \theta_i)}{\sigma_{PK} x^{\text{sim}}(t_{j_1}, \theta_i)}\right)^2 \right) \\
&+ \sum_{x_{i,j_2}^{\text{obs}} \in X_{i,PD}^{\text{obs}}} \left( -\frac{\log(2\pi)}{2} - \log\left(\sigma_{PD}\right) - \frac{1}{2}\left(\frac{x_{i,j_2}^{\text{obs}} - R_i(t_{j_2}, \theta_i)}{\sigma_{PD}}\right)^2 \right).
\end{aligned}$$

where $\theta_i = [\psi_i, \theta_{-\psi}]$. I have initially set just the parameter $V_c$ to vary between individuals, and then expanded the model to allow other parameters to vary.

Using the Chi package [4] for defining the mixed-effects model and Pints package [8] for optimisation and sampling algorithms, I performed parameter inference of the model with the simulated data. In this case performing maximum likelihood estimation would only provide estimates for individual-level parameters. However, estimating the population parameters and letting individual estimates be affected by the population values in the optimisation is important. Thus instead of optimising over the likelihood, I have determined the maximum-a-posteriori (MAP) estimate. It is easier in CHI to perform inference on the full PKPD problem, and thus for the mixed-effects inference I did not separate this into the PK and PD problems. The priors and transformations used for each parameter is listed in Table 5.1.

I ran the CMAES algorithm 10 times to determine the MAP with the results shown in Figure 5.3. The starting point of population-level parameters for run 1 was estimated using Non-compartmental Analysis (NCA) [18] on the pooled results, all other starting points were sampled from the prior. Some of the sampled starting points produced non-finite posterior scores, and for these runs the starting point was resampled until it had a finite posterior.

A

| | $S$, L/mg | $K_{cl}$, L/hr | $V_{c,typ}$, L | $\omega_{V_c}$ | $K_1$, L/hr | $V_1$, L | $R_0$, $10^3/\mu$L | $\gamma$ | MTT, hr | $\sigma_{PK}$ | $\sigma_{PD}$ | **Log-posterior** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Run 2** | 21.43 | 2.987 | 0.313 | 0.0 | 889.96 | 3.535 | 983.47 | 0.437 | 85.207 | 0.758 | 42.093 | -7944.83 |
| **Run 7** | 10.341 | 4.253 | 0.809 | 0.215 | 502.408 | 0.492 | 695.213 | 0.111 | 106.943 | 0.177 | 0.31 | -inf |
| **Conv.** | 20.257 | 2.819 | 0.693 | 0.314 | 2.342 | 4.951 | 983.334 | 0.437 | 85.069 | 0.088 | 41.881 | -2676.16 |
| **True** | 20.0 | 2.8 | 0.68 | 0.3 | 2.3 | 4.9 | 983.1 | 0.44 | 85.26 | 0.091 | 0.13 | - |

B

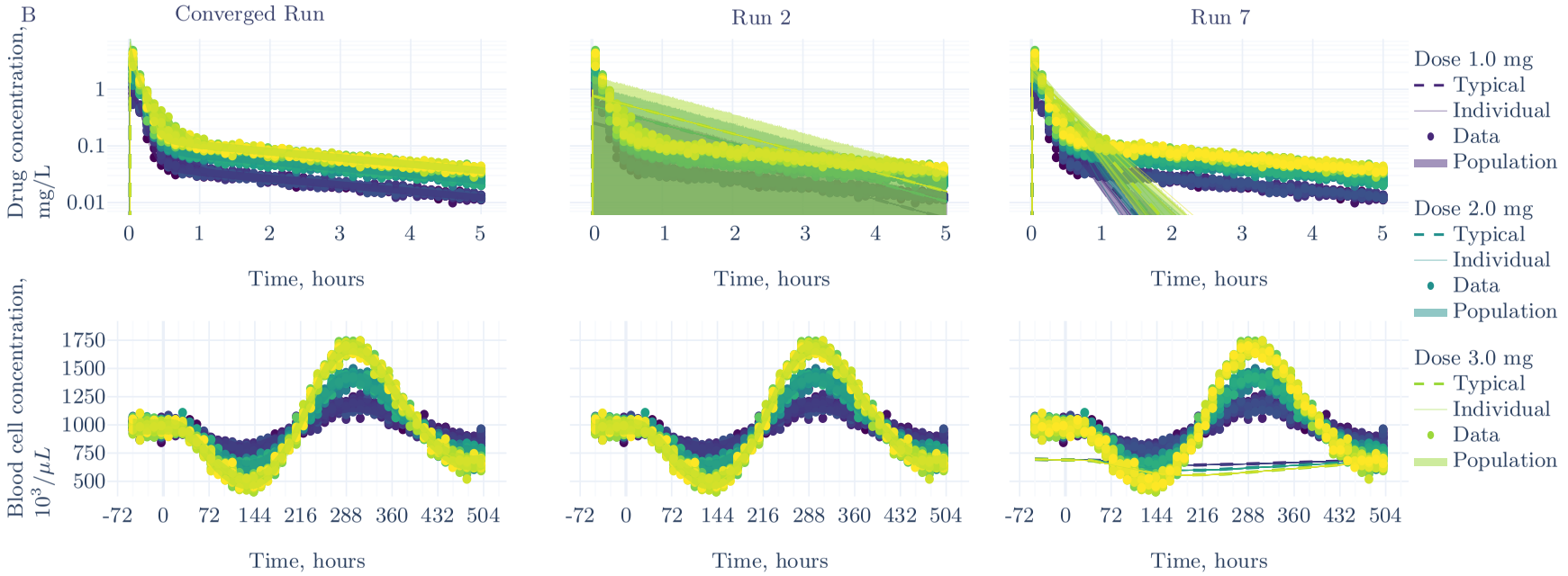



Figure 5.3: Results of Maximum-a-posterori (MAP) inference on the PKPD model with a mixed-effects $V_c$ parameter. A) Table of the maximum posterior estimate of each hyper-parameter and the "true" data-generating parameter values. The runs that had converged (Log-posterior value within 1.92 of maximum) have been condensed into a single row. B) The simulation of the output using the MAP, compared with the data. Run 1, run 2 and run 7 are shown, with run 1 representing the converged runs. The shaded region signifies the 5-th to 95-th percentile expected from the population model.

| Parameter | | Prior | Transformation |
|---|---|---|---|
| $K_{\text{cl}}$ | | Lognormal($K_{\text{cl},\approx}$, $0.75^2$) | $T(K_{\text{cl}}) = \log(K_{\text{cl}})$ |
| $K_1$ | | Lognormal($K_{\text{cl},\approx}$, $3^2$) | $T(K_1) = \log(K_1)$ |
| $V_c$ | $V_{c,i}$ | Lognormal($V_{c,typ}$, $\omega_{V_c}^2$) | $T(V_{c,i}) = \log(V_{c,i})$ |
| | $\log(V_{c,typ})$ | N($\log(V_{c,\approx})$, $0.3^2$) | None |
| | $\omega_{V_c}$ | Lognormal(0.3, $0.4^2$) | $T(\omega_{V_c}) = \log(\omega_{V_c})$ |
| $V_1$ | | Lognormal($V_{c,\approx}$, $0.6^2$) | $T(V_1) = \log(V_1)$ |
| $\sigma_{PK}$ | | Lognormal(0.1, $0.4^2$) | $T(\sigma_{PK}) = \log(\sigma_{PK})$ |
| $R_0$ | | Lognormal($R_{0,\approx}$, $0.5^2$) | $T(R_0) = \log(R_0)$ |
| MTT | | Lognormal($\text{MTT}_{\approx}$, $0.5^2$) | $T(\text{MTT}) = \log(\text{MTT})$ |
| $\gamma$ | | Lognormal(0.5, $1.2^2$) | $T(\gamma) = \log(\gamma)$ |
| $s$ | | Lognormal($s_{\approx}$, $0.5^2$) | $T(s) = \log(s)$ |
| $\sigma_{PD}$ | | Lognormal($0.1R_{0,\approx}$, $2.1^2$) | $T(\sigma_{PD}) = \log(\sigma_{PD})$ |

Table 5.1: Table of the priors and transformation used for each parameter. Approximations of some parameters, $\theta_k$, have been initially approximated using Non-compartmental analysis [18], and these are denoted as $\theta_{k,\approx}$. In the python package CHI, the log-normal mixed-effects model is parameterised by $\log(\theta_{typ})$ rather than $\theta_{typ}$ and so $V_{c,typ}$ has been pre-transformed during this analysis.

Using CMAES, the algorithm failed to find the maximum posterior value for two of the runs. As CMAES is a population based global optimiser, there is a possibility the population of points will not find the correct local maxima, and thus it is important to do multiple runs to check convergence. If no convergence occurs, it could be a sign that there is practical identifiability problems (see section 6).

To fully assess the performance of PINTS and CHI with potentially mismatched mixed-effects models, I initiated 10 maximum-a-posteriori runs each for four different assumptions and for each category 3 simulated data-set. These assumptions on the model were that:

- All parameters, excluding the noise parameter, are mixed-effects parameters. (True for the latter four data-sets)
- $V_c$ is the only mixed-effects parameter. (True for the first four data-sets)
- $K_{cl}$ is the only mixed-effects parameter.
- There are no mixed-effects parameters.

I used the CMAES optimiser and the starting points were generated in the same manner as above.

Results of the convergence of the optimisers on the eight simulated data-sets is shown in Figure 5.4. CMAES converged for all data-sets with only one mixed-effects

parameter. It did not easily converge when the data was generated with both mixed-effects parameters but only $V_c$ was assumed to be mixed-effects. We would usually expect CMAES to perform better with a simpler model than the more complicated model, however the cause of this difficulty in convergence is explored more in Chapter 6.1.

| | Assumptions | | | |
|---|---|---|---|---|
| **Dataset** | **All params M-E** | **$V_c$ M-E** | **$K_{cl}$ M-E** | **F-E** |
| 1 (1 M-E param) | 10 | 10 | 10 | 10 |
| 2 (1 M-E param) | 10 | 10 | 10 | 10 |
| 3 (1 M-E param) | 10 | 10 | 10 | 10 |
| 4 (1 M-E param) | 10 | 10 | 10 | 10 |
| 5 (2 M-E params) | 10 | 10 | 10 | 10 |
| 6 (2 M-E params) | 10 | 6, 3, 1 | 10 | 10 |
| 7 (2 M-E params) | 10 | 1, 8, 1 | 10 | 10 |
| 8 (2 M-E params) | 10 | 10 | 10 | 10 |

Figure 5.4: Number of runs, out of ten, that produced the similar log-likelihood score (within 1.98 of each other). Where multiple values are provided, these values are the number of runs that produced the highest, 2nd highest and 3rd highest log-likelihood scores respectively. This was performed on eight different data-sets generated with either one or two mixed-effect (M-E) parameters, and four different assumptions on the model (where F-E indicates the fixed-effects model).

## 5.4 Monolix

Monolix [39] is an application that determines individual and population parameters of PKPD models given some data and is a common tool used in current PKPD modelling. I ran Monolix on the one-compartment PK datasets as a comparison to the CHI-based workflow.

In monolix, population parameters are divided into three categories, fixed effects parameters ($\theta_{typ}$), random effect parameters ($\omega$) and residual error parameters. The model parameters are automatically set to be mixed effects (with both random and fixed effect parts) but can be changed to be fixed effect or fixed to a particular value. The mixed effect parameter distributions can also be chosen between normal, log-normal and logit-normal. Error models are chosen between constant, proportional and 2 variations of combined noise. Initial values of these parameters are all defaulted to the value one but can be changed by the user. Additionally, Monolix can

automatically find initial values for the fixed effects parameters. To do this, it uses a subset of the individuals, and optimises on the pooled data of these individuals. The optimisation is done by a "custom optimization method" and no other details on the algorithm is provided.

The first task that Monolix performs is to estimate the parameters using the SAEM (Stochastic Approximation Expectation-Maximization) algorithm. To use this algorithm, our problem needs to be redefined as a missing data problem. Let $g(\theta|Y)$ be the incomplete data likelihood, i.e. the likelihood of the parameters $\theta$ given the observed data, $Y$. And let $f(Z,\theta|Y)$ be the complete data likelihood, i.e. the likelihood of the parameters $\theta$ given all the data, $X=(Y,Z)$ where Z, the missing data, are the individual-level parameters, $\theta_{typ}+\omega\eta_i$, $\eta_i \sim N(0,1)$. Additionally, define $p(Z|Y,\theta)$ as the posterior distribution of the missing data given the the observed data Y (also known as the predictive distribution).

Expectation-Maximization (EM) is achieved through multiple iterations of maximising

$$Q(\theta|\theta') = \int_{\mathbb{R}^l} \log f(Z,\theta|Y)\, p(Z|Y,\theta')\, \mu dz,$$

where $\mu$ is a $\sigma$-finite positive Borel measure on $\mathbb{R}^l$ instead of directly maximising the log-likelihood [12]. I.e. at iteration $g$, $\theta_{g+1}$ is the value of $\theta$ that maximises $Q(\theta|\theta_g)$. In SAEM, $Q(\theta|\theta')$ is approximated and each iteration is broken down into simpler steps:

1. Simulation Step - Generate $m(g)$ realisations $Z_g(j)$ for $j=1,\ldots,m(g)$ of the missing data under posterior density $p(Z|Y,\theta_k)$.

2. Stochastic approximation step - Let

$$\hat{Q}_g(\theta) = \hat{Q}_{g-1}(\theta) + \gamma_g \left( m(G)^{-1} \sum_{j=1}^{m(g)} \log f(Z_g(j),\theta) - \hat{Q}_g(\theta) \right)$$

   where $\gamma_{g_{g\geq 1}}$ is a sequence of step sizes.

3. Maximisation step - Let $\theta_{g+1}$ be the value of $\theta$ that maximises $\hat{Q}_g(\theta)$.

In this algorithm there are two optimisation parameters at each iteration, the stepsize, $\gamma_g$, and the number of simulations, $m(g)$, which need to be chosen. Monolix sets $\gamma_g = 1$ for $1 \leq k \leq G$ and $\gamma_g = \frac{1}{g-G}$ for $g > G$ and $m(g) = L$, where $G = 500$ and $L = 3$ by default.

In Non-linear mixed effects problems, there are additional difficulties in the simulation step as the posterior distribution cannot be directly sampled from. In Monolix, the Hastings-Metropolis MCMC algorithm is used to sample from this distribution with $L$ chains. Monolix also utilises simulated annealing to keep the area of exploration wide at the start and discourage chains getting stuck in local minima. This is achieved by limiting the decrease in $\omega$ to 5%.

SAEM does not work for fixed-effects parameters that do not have variability as there is no posterior density, $p\left(Z|Y,\theta_k\right)$ to sample from in the simulation step. By default the Nelder-Mead simplex algorithm is used to optimise these parameters. Optionally, artificial variability could be added for these parameters, allowing SAEM to be used instead, and then the variance is progressively reduced until it reaches 1e-5.

Once the population parameters are acquired, the individual-level parameters can be estimated using Empirical Bayes Estimates (EBEs). The EBEs are the mode of the conditional parameter distribution, $p\left(Z_i|Y_i,\hat{\theta}\right)$ where $Z_i$ are the individual-level parameters of individual $i$, $Y_i$ is the observed data of individual $i$, and $\hat{\theta}$ are the optimised population parameters from the previous step. This problem is redefined using Bayes law and the maxima of the distribution, $p\left(Y_i|Z_i,\hat{\theta}\right)p\left(Z_i|\hat{\theta}\right)$ is found using the Nelder-Mead Simplex algorithm. Optionally, the full conditional distribution can also be obtained using MCMC sampling.

I compared the results produced by CHI on the category 3 data to results gathered from Monolix. I intiated 4 SAEM runs each for the four different assumptions and for each simulated data-set. The starting point for run 1 was the same as used in run 1 for CMAES (see section 5.3). The starting points for runs 2 and 3 were auto-generated using Monolix's auto-initialise tool. And the starting point for run 4 was set to the default when creating a new model.

Monolix had difficulty converging to a reasonable result, when any fixed-effects parameters were not allowed inter-individual variability during the optimisation (i.e. Nelder-Mead simplex was used instead of SAEM). Monolix produced the error that the optimiser failed to meet the convergence criteria, unless all parameters were set to fixed-effects. In this latter case the optimiser met the convergence criteria very quickly and returned parameter values between 1.5 and 2 times the data-generating values. I moved to using SAEM with variability during the exploratory phase that is forcibly reduced towards zero and then Nelder-Mead simplex during the final stage. In this case the convergence error occurred for the fully fixed-effects model but did not for the other models.

A Monolix run was quicker than a CHI/PINTS run. However only one run was able to be initialised at a time and setting up each run manually significantly slowed down the process. Additionally, it was not easy to compare results between different runs within the application.

## 5.5 Conclusion

In this chapter I used the data as described in Chapter 4 to find parameters for the models described in Chapter 2 by optimising an objective function. This objective function was either the Likelihood function, $P\left(X^{\text{obs}}|\theta\right)$, to find the maximum likelihood estimate (MLE) or the posterior function, $P\left(\theta|X^{\text{obs}}\right)$, to find the Maximum-a-posteriori (MAP) estimate. In conclusion, the MLE could not identify the data-generating parameters of the noise model for the fixed-effects simulated data, which could indicate identifiability issues with these parameters. Running multiple runs of the MAP estimation in mixed-effects models shows that the optimiser does not always find the optimum, thus it is important to do multiple runs and check convergence of MAP estimates. Doing a similar analysis in the program Monolix can be difficult, as it is generally designed towards the method of starting with the most complex population model rather than the least. Additionally, Monolix is not set up to do multiple runs with different start points. However, the MLE of the fixed effects model and the MAP of the mixed-effects model has been found. With these we can do further analysis to identify the potential cause of these problems. Overall these methods are quick but they do not provide any information on the uncertainty in these parameter estimates.

# Chapter 6

# Practical identifiability

As seen in section 5, a model that is structurally identifiable can still pose difficulties for parameter inference. This may be due to practically non-identifiable parameters. Practical non-identifiability occurs when different parameter values cannot be discerned from noisy data. If a model is structurally non-identifiable then it is practically non-identifiable. In this section I will do a review of the methods currently available for profile likelihoods, analyse the PKPD model and data using Profile likelihoods, and increase the efficiency of these methods.

To determine the practical identifiability, the confidence in parameter estimates must be quantified. One way of performing this to use the Fisher Information matrix (FIM),

$$I\left(\hat{\theta}\right) = -\frac{\partial^2}{\partial\theta^2}\log\left(\mathcal{L}_Y\left(\theta\right)\right),$$

where $\mathcal{L}_Y\left(\theta\right)$ is the log-likelihood of the parameter $\theta$, given data $Y$, to determine the standard errors of the parameters. The inverse of this matrix, $C\left(\hat{\theta}\right) = I^{-1}\left(\hat{\theta}\right)$, is an approximation to the covariance matrix of the parameters. Thus the square root of the diagonals of this matrix give the standard errors, $s.e\left(\hat{\theta}_k\right) = \sqrt{C_{kk}\left(\hat{\theta}\right)}$ and the correlation between parameters is determined as $corr\left(\theta_{k_1}, \theta_{k_2}\right) = \frac{C_{k_1k_2}}{s.e\left(\theta_{k_1}\right)s.e\left(\theta_{k_2}\right)}$. The confidence intervals are then determined from these standard errors. This method is commonly used to determine identifiability especially for PKPD models [54, 56] and can be calculated via the Monolix program. However, this method assumes the log-likelihood is exactly quadratic and is inaccurate when the parameters are not [43], and, in non-linear PKPD models it is unlikely for all parameters to be exactly quadratic.

An alternative method to determining the practical identifiability of a model, that is starting to be used more regularly, is the profile-likelihood method [55, 14]. In this

method, each parameter is varied around the MLE value, and the log-likelihood for the other parameters re-optimised, i.e.,

$$PLL_k\left(\theta_k\right) = \max_{\theta_{-k}} \left(\log\left(P\left(X^{\text{obs}}|\theta_k, \theta_{-k}\right)\right)\right), \tag{6.1}$$

for each parameter $\theta_k$, where $\theta_{-k}$ are all parameters excluding $\theta_k$. The confidence interval is then determined from this profile likelihood. The interval is the range of $\theta_k$ that

$$PLL_k\left(\theta_k\right) > LL\left(\theta^{(\text{MLE})}\right) - \frac{q_1\left(1-\alpha\right)}{2},$$

where $LL\left(\theta^{(\text{MLE})}\right)$ is the log-likelihood of the maximum likelihood estimate, and $q_1\left(1-\alpha\right)$ is the $(1-\alpha)$th quantile of the $\chi^2$ distribution with 1 degree of freedom. For a 95% confidence interval, $q_1\left(1-\alpha\right) = 3.841$. This method produces more accurate confidence intervals, is invariant under transformations of the parameters and can be asymmetric around the MLE, unlike the FIM derived confidence intervals. [64]

In some papers, a parameter is considered unidentifiable if these confidence intervals are wide. This determination is then subject to the researchers choice in what wide constitutes [55]. In Raue *et. al.*'s paper on identifiability [45], they more formally defined this requirement. If the confidence intervals from the profile likelihood, $PLL_k\left(\theta_k\right)$, is infinite then $\theta_k$ is practically unidentifiable. Also, if there exists no value of $\alpha$ at which the confidence interval is finite then $\theta_k$ is also locally structurally non-identifiable, where even with infinite noiseless data the confidence interval will always be infinite [64]. This formal definition cannot apply to the FIM confidence intervals as they are always finite and an unidentifiable parameter as per Raue *et. al.*'s definition will be poorly estimated by the FIM, except in the case that the profile likelihood is completely flat.

The profile likelihood can also be extended into higher dimensions. For example, a bivariate profile likelihood would be the surface defined by the equation,

$$PLL_{k_1,k_2}\left(\theta_{k_1}, \theta_{k_2}\right) = \max_{\theta_{-k_1,-k_2}} \left(\log\left(P\left(X^{\text{obs}}|\theta_{k_1}, \theta_{k_2}, \theta_{-k_1,-k_2}\right)\right)\right),$$

for each pair of parameters, $\theta_{k_1}, \theta_{k_2}$. As well as an indication of identifiability this provides information of parameter interdependence. [6]

If parameters are unidentifiable then there are several ways to change the problem such that it is identifiable. If there are only practical identifiablity problems, i.e there would be no identifiablity problem for infinite, noiseless data, then gathering more data may help. These new experiments can be designed using the model, in order to maximise identifiabilty and minimise confidence intervals [37, 38, 20]. If it is not

feasible to do this or there are structural identifiability issues, then a change to the model can be made to reduce the parameter space by reparamatisation, keeping the unidentifiable parameter at a fixed value, or introducing an equation to relate two unidentifiable parameters, as informed by the bivariate profile likelihoods [64, 6].

The calculation for profile-likelihoods can be difficult. The brute force option to find the confidence interval is to choose values of $\theta_k$ at regular intervals in some interval $I \subset \Theta_k$, optimising Equation 6.1 at each value of $\theta_k$. This method can take a long time depending on how quickly the optimisation can be done and how many points are selected along the interval but provides the shape of the profile likelihood. Another method proposed by Venzon and Moolgavkar [60] is to use a modified Newton-Raphson algorithm to solve the system of equations

$$LL(\theta) - LL\left(\hat{\theta}\right) + \frac{q_1(1-\alpha)}{2} = 0, \tag{6.2}$$

$$\frac{\partial L}{\partial \theta_\kappa}(\theta) = 0, \forall \kappa \neq k, \tag{6.3}$$

for $\theta_k \in \theta$ where $\hat{\theta}$ is the MLE. This is an equivalent problem as the first part provides the end points of the C-I, and the second part ensures the optimum in the profile likelihood. This modified Newton-Raphson method utilises a quadratic approximation to $LL(\theta)$ to inform the step size, however calculating the partial derivatives is still costly.

To acquire the profile likelihhods, I have decided initially to use the brute force method, but utilising a quick local optimiser instead of a global optimiser. To ensure accuracy and efficiency in the optimiser, I will start at the MLE and work out. At each step I will use the previously optimised values as the start point for the local optimiser.

Acquiring the profile likelihoods for the fixed-effects simulated data (category 1) confirms that the non-identifiability of this model stems from the noise parameters. This can be seen in Figure 6.1 where the profile likelihoods of the PD model parameters all have clear intersections with the horizontal line at $LL\left(\hat{\theta}\right) + \frac{q_1(1-\alpha)}{2}$, whereas the only noise parameter that is identifiable is $\eta$. There is also a shelf to the left of the maximum and this may affect the MLE results, however, this shelf is likely to be an artefact of the imperfect optimisation as it is not smooth. The bivariate profile likelihoods of these parameters also show the inter-parameter interactions.

Performing the same analysis on the Experimental data, I acquired the profile likelihoods of the PD parameters when combined noise is assumed. These results are shown in Figure 6.2. The profile likelihoods show that the PD model parameters,

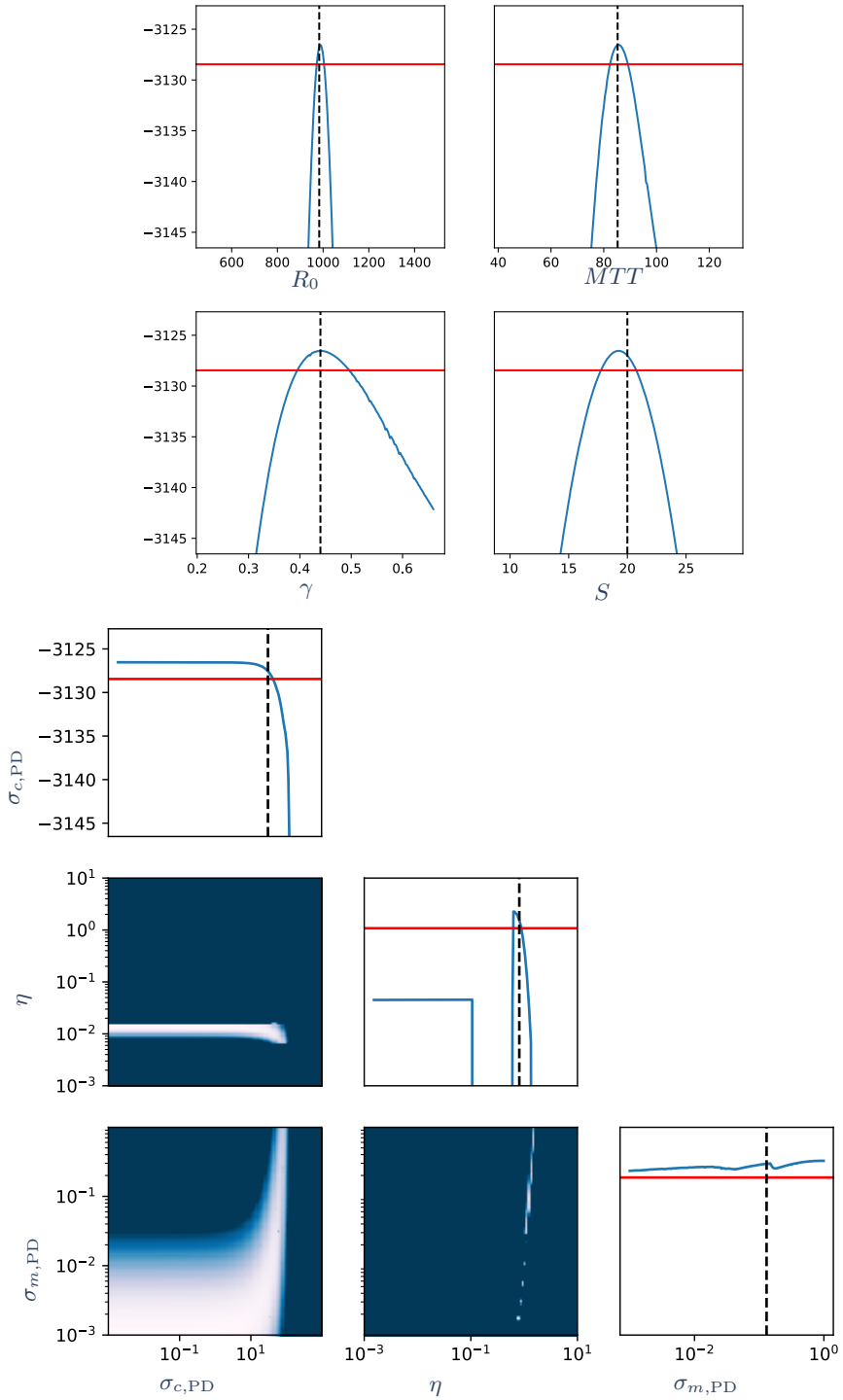



Figure 6.1: Profile Likelihoods of the PD parameters on category 1 simulated data, assuming combined noise and fixed-effects model. Red confidence line is at $\frac{q_1(0.95)}{2} = 1.92$ below the log-likelihood of the maximum likelihood estimate, and black dashed line shows the parameter values used to simulate the data. (A) The profile likelihood of the Friberg model parameters. All these parameters show identifiability. (B) The profile likelihoods and bivariate profile likelihoods of the combined noise model parameters.

($R_0$, MTT, $s$ and $\gamma$), are identifiable, with combined noise. However, $\sigma_{c,\mathrm{PD}}$ is the only noise parameter that is identifiable. I further investigated the identifiability of the noise model by fixing various parameters to determine the conditions under which the parameters are identifiable. The results from this are shown in Figure 6.3A.

Fixing $\sigma_{c,\mathrm{PD}}$ to zero (i.e. converting the noise to multiplicative noise) ensures that $\eta$ is identifiable but $\sigma_{m,\mathrm{PD}}$ is against the edge of the parameter space. Fixing $\eta$ to 1 means that $\sigma_{c,\mathrm{PD}}$ is identifiable but $\sigma_{m,\mathrm{PD}}$ is still not identifiable. Fixing $\sigma_{m,\mathrm{PD}}$ to zero (i.e. converting the noise to constant noise) ensures that $\sigma_{m,\mathrm{PD}}$ is identifiable but $\eta$ is structurally unidentifiable as the noise from the multiplicative part, $\sigma_m x^{\mathrm{sim}}(t_j)^{\eta}\epsilon_j$, is always zero. I also fixed pairs of parameters as shown in Figure 6.3B. Initially, $\sigma_{m,\mathrm{PD}}$ was fixed to zero and $\eta$ fixed to 1 which produced the same results for $\sigma_{c,\mathrm{PD}}$ as when just $\sigma_{m,\mathrm{PD}}$ was fixed. I could not set both $\sigma_{c,\mathrm{PD}}$ and $\sigma_{m,\mathrm{PD}}$ to zero so instead I fixed $\sigma_{c,\mathrm{PD}}$ to 193.08 (the point where maximum likelihood occurred) and $\sigma_{m,\mathrm{PD}}$ to 0.01 to acquire the profile likelihood of $\eta$. This did not produce an identifiable $\eta$. Finally I set both $\sigma_{c,\mathrm{PD}}$ to zero and $\eta$ to 1 and this did produce an identifiable $\sigma_{m,\mathrm{PD}}$.

In these graphs there seem to be some discontinuities. These are likely due to inaccuracies in the numerical calculations rather than the profile likelihood, which tends to be smooth. This potentially comes from either the ODE solver or the optimiser. In the ODE solver, as the parameters and solution changes, the solver can choose a different set of time points to solve the the ODE at and interpolate between [9]. This can cause a discontinuous change in the numerical solution of the ODEs and thus the log-likelihood . In this case any jumps in the solution should be small as the errors in calculating the ODE should be limited by the solver. With the optimiser, to maintain efficiency, I have used a local optimiser. I would expect the maximising parameters, $\theta_{-k}$, to follow along a path through the parameter space and thus the local optimiser, starting at the previous point, would be able to find this parameter. However, if there are multiple local optima, then our algorithm may only follow the path of one local optimum whose log-likelihood decreases at a quicker rate than another optimum. This second local optimum could then become the global optimum. If then, the current local maximum merges into the global maximum, this would be seen in these graphs as a sudden jump upwards. This latter scenario may be the explanation of the separated shelves seen in Figure 6.1B for $\eta$ and 6.2A for $\sigma_{c,PD}$ with fixed $\eta$. In either case, the effect on the identifiablity results, and confidence intervals are not affected.

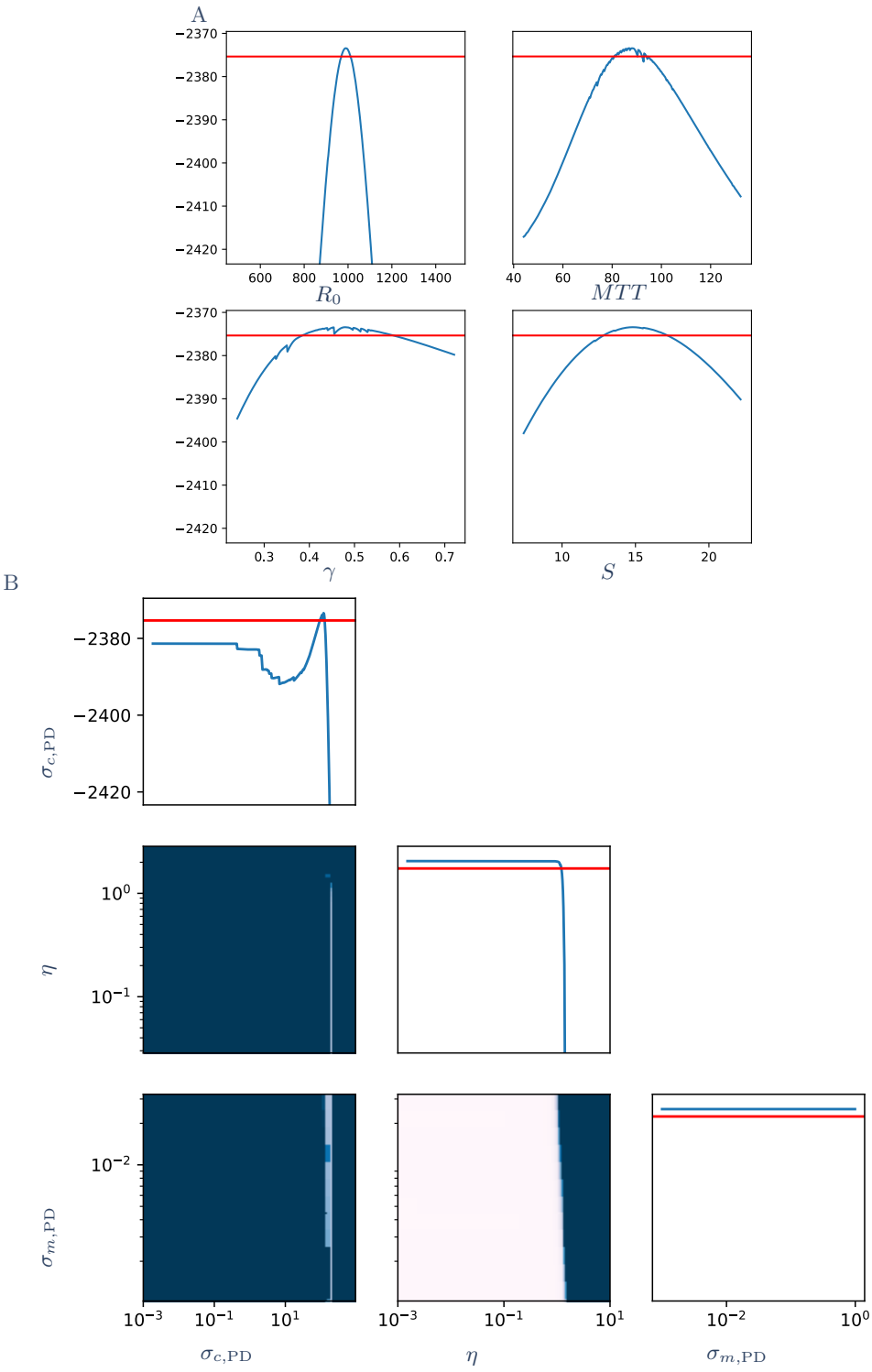



Figure 6.2: Profile likelihoods of the PD parameters on experimental data, assuming combined noise and fixed-effects model. Red line is at 1.92 below the maximum of the log-likelihood. (A) The profile likelihood of the Friberg model parameters. All these parameters show identifiability. (B) The profile likelihoods and bivariate profile likelihoods of the combined noise model parameters.

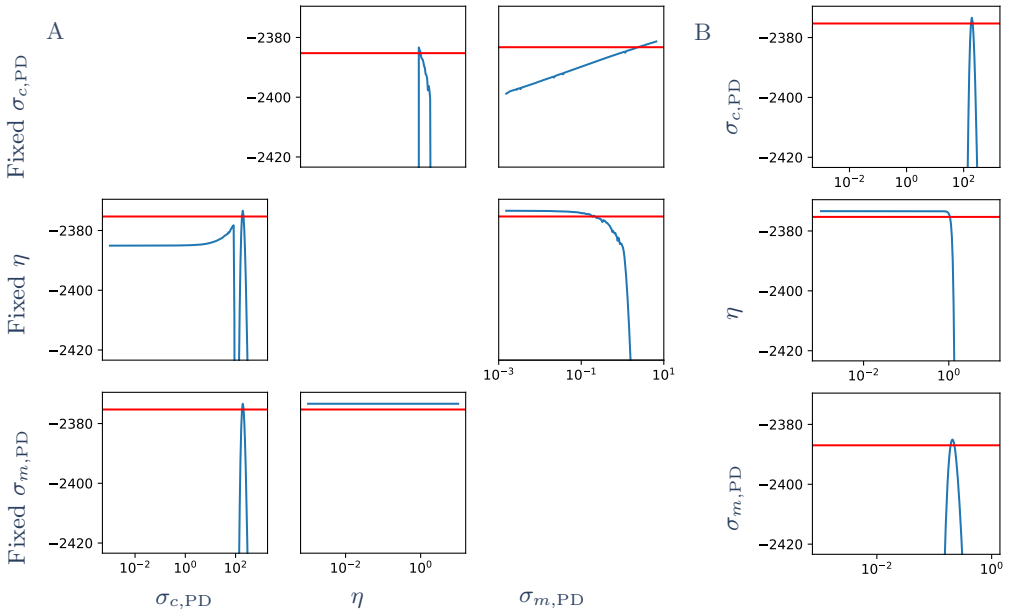



Figure 6.3: Profile likelihoods of the PD Noise parameters while some parameters are fixed to a value. Red line is at 1.92 below the maximum of the log-likelihood. A) Single parameters fixed. From top to bottom: $\sigma_{c,\mathrm{PD}}$ is fixed to zero, $\eta$ is fixed to one, and $\sigma_{m,\mathrm{PD}}$ is fixed to zero. B) Two parameters are fixed. From top to bottom: $\sigma_{m,\mathrm{PD}}$ is fixed to zero, $\eta$ to one; $\sigma_{c,\mathrm{PD}}$ is fixed to 193.08, $\sigma_{m,\mathrm{PD}}$ to 0.01;and $\sigma_{c,\mathrm{PD}}$ is fixed to zero, $\eta$ to one.

Acquiring the profile-likelihood for mixed-effects models is more complicated. The likelihood as defined by

$$P\left(X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}|\theta_1,\ldots,\theta_{n_i}\right)=\prod_{i=1}^{n_i}P\left(X_i^{\text{obs}}|\theta_i\right)$$

does not incorporate the hyper-parameters. However, you could consider the likelihood of the hyper-parameters to be the posterior without the hyper-prior term, i.e,

$$\begin{aligned}LL\left(\theta_{\text{typ}},\Omega\right)=&P\left(\theta_1,\ldots,\theta_{n_i},X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}|\theta_{\text{typ}},\Omega\right)\\=&P\left(\theta_1,\ldots,\theta_{n_i},|\theta_{\text{typ}},\Omega\right)\cdot P\left(X_1^{\text{obs}},\ldots,X_{n_i}^{\text{obs}}|\theta_1,\ldots,\theta_{n_i}\right),\end{aligned}$$

for individual-level parameters $\theta_1,\ldots,\theta_{n_i}\in\Theta$. This is a similar approach to how Monolix finds the MLE. To find the profile likelihood of parameter $\theta_k\in\{\psi_{typ},\theta_{-\psi},\Omega\}$, this likelihood is optimised over the individual-level parameters and other hyper-parameters.

The second issue with acquiring the profile likelihoods on mixed effects problems is the time efficiency. For the fixed effects problem, it took several hours to find the single and bivariate profile likelihoods for all parameters in one model. With multiple models that have at least 46 (45 individuals and a variability parameter for each mixed-effects parameter) extra parameters, the computational expense grows dramatically. As I am more concerned with population dynamics in this case study, I can exclude the profile likelihoods of individual-level parameters from the analysis. However, these parameters are still optimised over and thus will slow down the optimisation in the analysis of the hyper-parameters.

In order to create a more efficient process of finding the profile likelihoods, I acquired the profile likelihood curves for the simple one compartment PK mixed-effect model as it has the fewest parameters and thus should be easy to repeat analysis on. The profile likelihood for each simulated data-set in category 3 and assumption as detailed in Chapter 5.3 was calculated. The profile likelihood curves for the parameters $K_{cl}$ and $\omega_{K_{cl}}$ are shown in figure 6.4. $\omega_{K_{cl}}$ is unidentifiable for data-sets 1-4 (generated with $\psi=[V_c]$), under the modelling assumption that $\psi=[V_c,K_{cl}]$ . In this case, the profile log-likelihood increases as the parameter tends towards zero. This behaviour is expected, as these data-sets were generated with $\omega_{K_{cl}}=0$, and the behaviour disappears for data-sets 5-8, where $\omega_{K_{cl}}\neq 0$. The Monolix runs estimated values close to zero, however, the MAP estimate was greater as the hyper-prior had zero probability at $\omega_{K_{cl}}=0$. This behaviour also disappears when only $K_{cl}$ is assumed

to be mixed-effects in the modelling. In this case, although the data-generating value of $\omega_{K_{cl}}$ is zero, some variability in $K_{cl}$ is optimised to compensate for the lack of variability in $V_c$.

When $V_c$ only is allowed to vary there is a lack of smoothness in the profile log-likelihood of $K_{cl}$ for all data-sets. In these cases it still leads to a single finite confidence interval, and thus is identifiable by definition. However the multi-modal nature of this does not lead to a single maximum likelihood estimate within the confidence interval and may be the cause of the failure to converge in the above cases, if it is not an artefact of the optimisation problem.

The profile likelihood curves for the parameters $V_c$ and $\omega_{V_c}$ are shown in figure 6.5. In these graphs the jaggedness are present again, but only when $K_{cl}$ is allowed to vary and not $V_c$. The jaggedness also leads to practically unidentifiable parameters for some of the data-sets, where the profile likelihood curve intersects the confidence interval line multiple times. It also seems that, although converged to the same answer, both PINTS and Monolix optimised to the lower local optimum rather than the global optimum for data-sets 1 to 4.

## 6.1 Investigating methodology on Mixed-Effects models

To further investigate the cause of these discontinuities I acquired further detail for data-set 5. I looked at the individual contributions to the overall likelihood in the likelihood profile curve,

$$\log L\left(\theta_i, \theta_{-k,\text{typ}}, \theta_\text{k}, \Omega | X_\text{i}^\text{obs}\right) = LL_i\left(\theta_\text{i}\right) + \log P\left(\theta_i, |\theta_{-k,\text{typ}}, \theta_\text{k}, \Omega\right),$$

where $LL_i\left(\theta_\text{i}\right)$ is the individual Log-Likelihood and $(\theta_i, \theta_{-k,\text{typ}}, \Omega)$ are the optimised parameter values as $\theta_\text{k}$ is varied. I also investigated how the individual-level parameters, $\theta_\text{i}$, change as $\theta_\text{k}$ varies. The results are shown in Figure 6.6A.

As can be seen from the graphs, the optimised $\theta_\text{i}$ and $\theta_{-k,\text{typ}}$ values are not smooth. They stay stationary for a while and then suddenly all change at the same point, in a step-like function. These step changes correspond to small jumps in the individual contribution curves which sum up into the larger jumps in the overall profile likelihood.

This problem may be related to the Nelder-Mead simplex algorithm, implemented by PINTS, and it's failure to optimise. Thus I compared this with how well CMA-ES performs as the optimiser (instead of Nelder-Mead) when obtaining the profile likelihood, as shown in Figure 6.6B. When CMA-ES is used, the parameters are

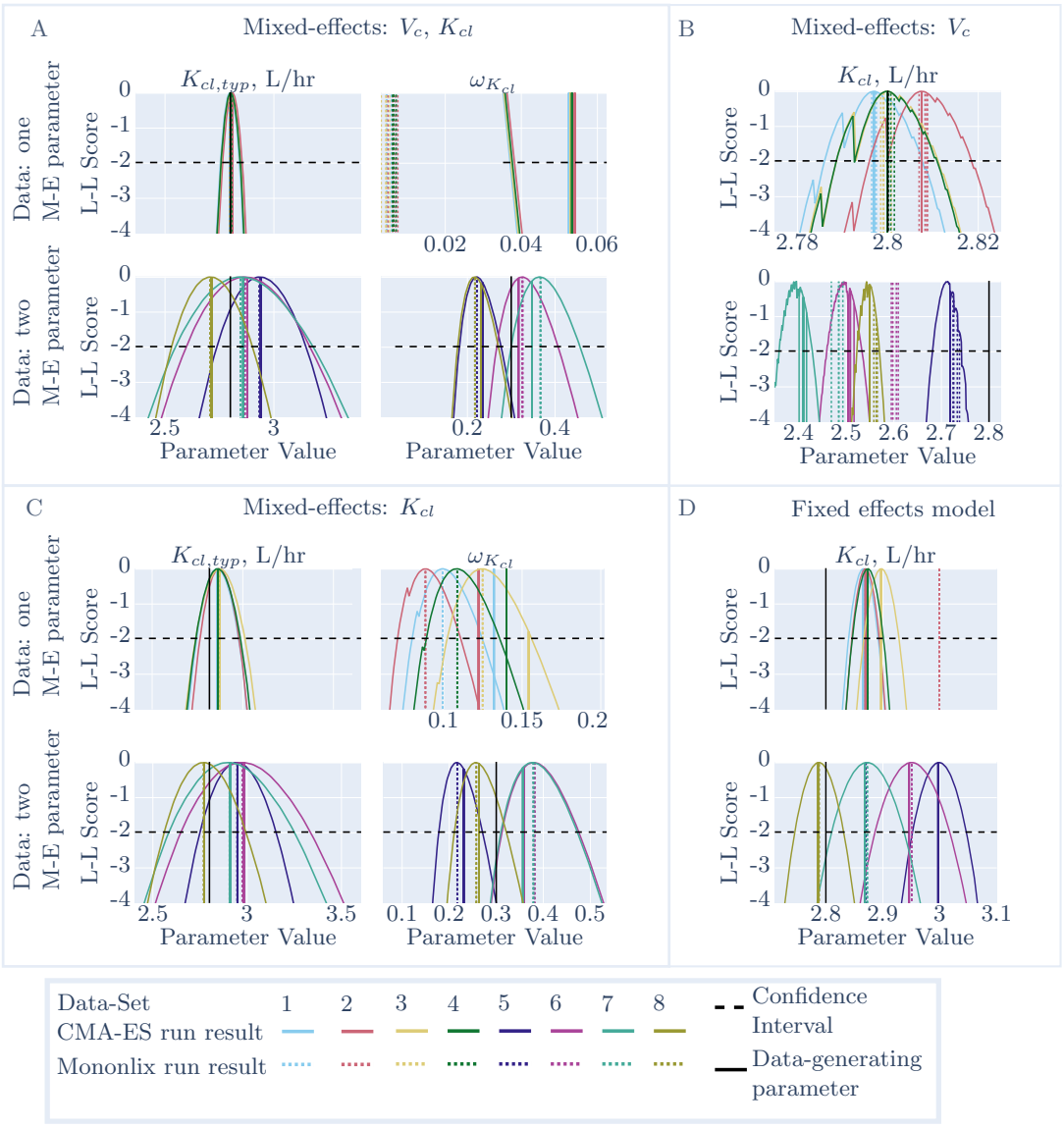



Figure 6.4: Profile likelihood curves for $K_{cl}$, and $\omega_{K_{cl}}$ (when included in the model), under the four assumptions. A) Both parameters $V_c$ and $K_{cl}$ are assumed to be mixed-effects, B) Only parameter $V_c$ is mixed-effects, C) Only parameter $K_{cl}$ is mixed-effects, D) No parameters are mixed-effects. The top graphs in each assumption is for Data-sets 1 to 4, where only $V_c$ was mixed-effects in generating the data. The bottom graphs in each assumption is for Data-sets 5 to 8, where both parameters were mixed-effects when generating the data. Graphs have been normalised such that the maximum log-likelihood is at zero. Vertical solid lines show where each run of the PINTS implemented CMA-ES method optimised to. Vertical dashed lines show where each run of the Monolix implemented SAEM method optimised to. Black vertical line is the data-generating values. Horizontal dashed line is 1.92 below the maximum and shows the confidence intervals.

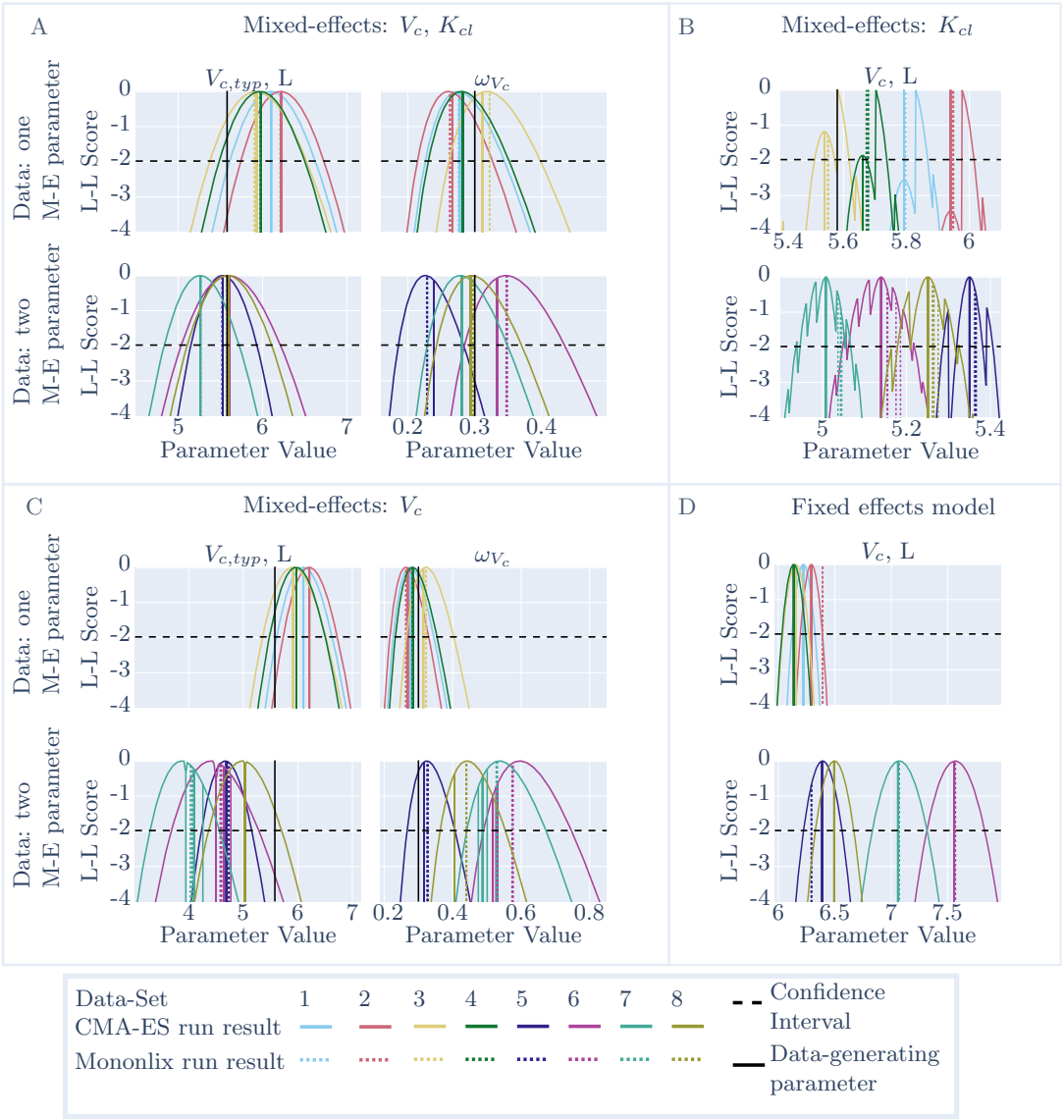



Figure 6.5: Profile likelihood curves for $V_c$, and $\omega_{V_c}$ (when included in the model), under the four assumptions. A) Both parameters $V_c$ and $K_{cl}$ are assumed to be mixed-effects, B) Only parameter $K_{cl}$ is mixed-effects, C) Only parameter $V_c$ is mixed-effects, D) No parameters are mixed-effects. The top graphs in each assumption is for Data-sets 1 to 4, where only $V_c$ was mixed-effects in generating the data. The bottom graphs in each assumption is for Data-sets 5 to 8, where both parameters were mixed-effects when generating the data. Graphs have been normalised such that the maximum log-likelihood is at zero. Vertical solid lines show where each run of the PINTS implemented CMA-ES method optimised to. Vertical dashed lines show where each run of the Monolix implemented SAEM method optimised to. Black vertical line is the data-generating values. Horizontal dashed line is 1.92 below the maximum and shows the confidence intervals.

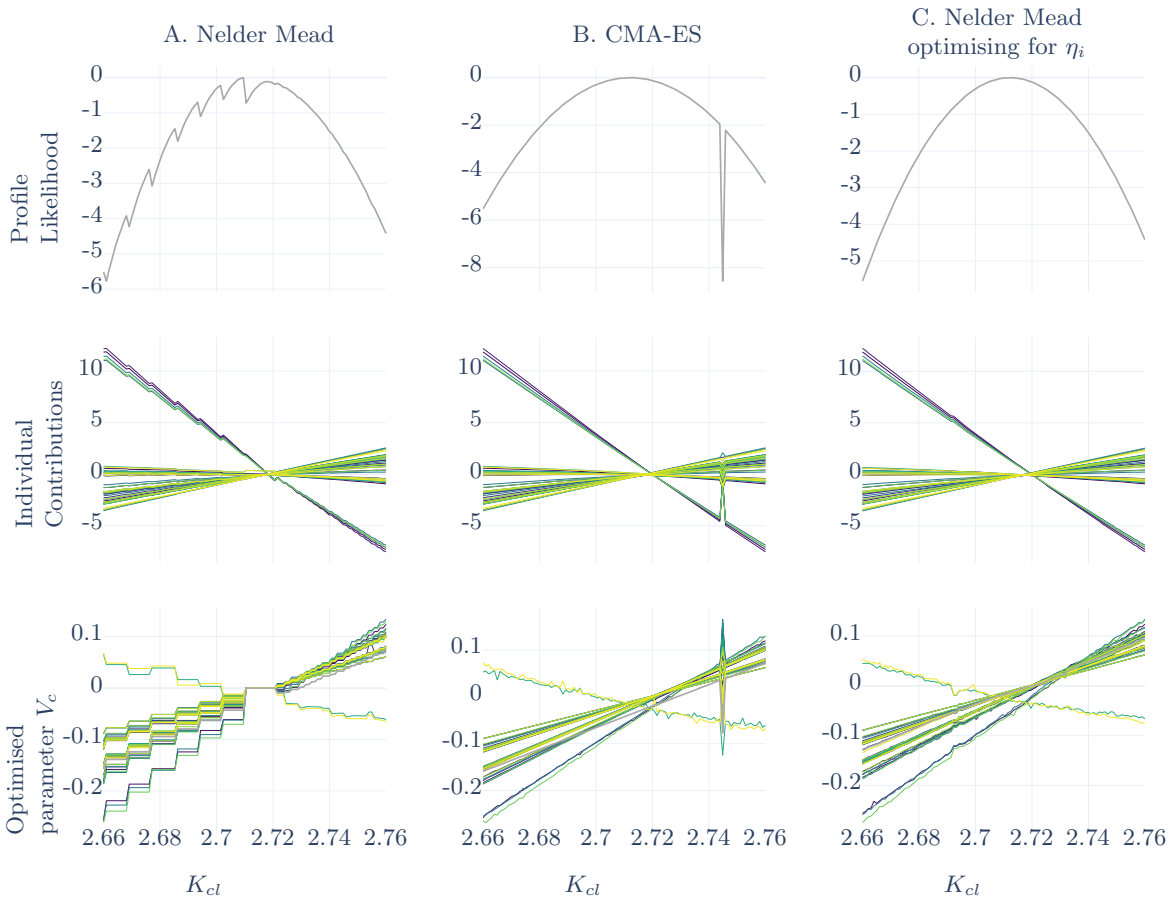



Figure 6.6: Profile likelihood graphs (top), change in individual log-likelihoods (middle) normalised by the same value at the MLE, and change in $V_{c,i}$ (bottom) compared with the MLE. These were produced for parameter $K_{cl}$ with Data-set 5 (two mixed effects parameters) under the assumption that $V_c$ only is mixed effects. A comparison between the standard Nelder-Mead (A), CMA-ES (B), and Nelder-Mead finding the random effects $\eta_i$ instead of individual-level parameters (C) is made.

continuously optimised which results in a smoother curve excluding a single point where the optimiser found an optimum further away from the initial point. The problem, however, with using CMA-ES in this fashion is speed. Nelder-mead took 13 minutes to produce one of the graphs from Figure 6.6, whereas CMA-ES took approximately 5 hours to produce the same graph. This makes it unreasonable to use when wanting multiple profile likelihoods, including bivariate profiles, on a more complicated system.

There is also the possibility that Nelder-Mead is having difficulty due to the strong correlation between $V_{c,i}$ and $V_{c,\text{typ}}$ for $1 \leq i \leq n_i$. If the system is in the case where individual-level parameters are in an optimal position in relation to the prior (which is where the optimiser starts off at), i.e. $\log P\left(V_{c,i}, |V_{c,\text{typ}}, \omega_{V_c}\right)$ is maximised for a set $\{V_{c,i} | 1 \leq i \leq n_i\}$, then any change to $V_{c,\text{typ}}$ will reduce the profile likelihood. If instead only one $V_{c,i}$ is changed, then the individual prior will reduce but $LL_i\left(\theta_\text{i}\right)$ may increase, thus the overall likelihood will only improve if the individual likelihood is a great enough improvement (which can be seen as the jumps in the profile likelihood). If a step is made in the right direction ($V_{c,\text{typ}}$ increases along with all the $V_{c,i}$) then convergence could be found. Nelder-Mead does not have enough information to make this choice, but a gradient-based optimiser would. Unfortunately, the gradient for this system is expensive to calculate. Another solution is to transform the optimisation space, such that $\eta_i$ is optimised instead of $V_{c,i}$, where $V_{c,i} = V_{c,typ} e^{\omega_{V_c} \eta_i}$, $\eta_i \sim N\left(0, 1\right)$. The results from optimising over $\eta_i$ is shown in Figure 6.6C. This shows that the method works for this example. However, it still fails to provide a smooth curve for all data-sets as shown in Figure 6.7A.

This implies that Nelder-Mead may not be an appropriate optimiser to use for profile likelihoods. I then investigated a selection of local optimisers provided by Scipy (which has a larger collection of local optimisers than PINTS). I ran these optimisers against each other and timed them. The most efficient optimiser was the Powell method which took approximately 10 minutes to calculate the profile likelihood for $K_{cl}$. It was slightly quicker than the PINTS implementation of Nelder-Mead and produced a smoother curve. The Nelder-Mead implementation in Scipy also produced a smoother curve but at a much greater cost of time. This is likely due to the termination criteria allowing Scipy to run more iterations.

The Powell method is a conjugate direction method. This method optimises the objective function by optimising along $N_p$ search vectors, $\left\{s_1, \ldots, s_{N_p}\right\}$, $s_i \in \Theta$, sequentially. This changes the optimisation problem from an $N_p$-dimensional problem

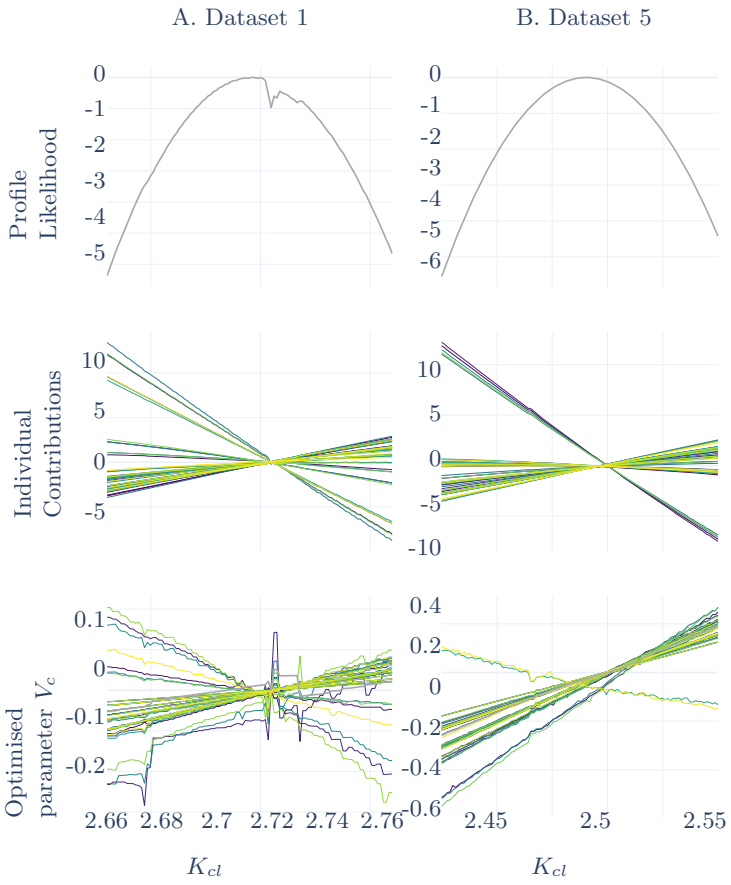



Figure 6.7: Profile likelihood graphs (top), change in individual log-likelihoods (middle) normalised by the same value at the MLE, and change in $V_{c,i}$ (bottom) compared with the MLE. These were produced for parameter $K_{cl}$ under the assumption that $V_c$ only is mixed effects. Nelder-Mead optimiser was used and the random effects $\eta_i$ instead of individual-level parameters were optimised. Results are acquired for (A) Dataset 1 (generated with one mixed-effects parameter) and (B) Data-set 5 (two mixed effects parameters)

A

| | Method | Powell | | | Nelder-Mead in PINTS | | | Nelder-Mead in Scipy | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Extrap. | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| TIME | $V_c$ | 12.79 | 12.65 | 11.31 | 3.98 | 3.95 | 3.98 | 38.08 | 38.16 | 38.09 |
| | $K_{cl}$ | 10.19 | 9.95 | 10.02 | 11.82 | 11.64 | 11.65 | 41.33 | 41.3 | 40.77 |
| | Total | 22.98 | 22.6 | 21.33 | 15.80 | 15.59 | 15.63 | 79.46 | 79.47 | 78.8 |
| ERROR | $V_c$ | 9.3 | 9.3 | 9.3 | 93.99 | 93.99 | 93.99 | 22.69 | 22.69 | 22.69 |
| | $K_{cl}$ | 0.3 | 0.3 | 0.3 | 101.15 | 101.15 | 101.15 | 0.32 | 0.32 | 0.32 |
| | Total | 9.6 | 9.6 | 9.6 | 195.14 | 195.14 | 195.14 | 23.01 | 23.01 | 23.01 |

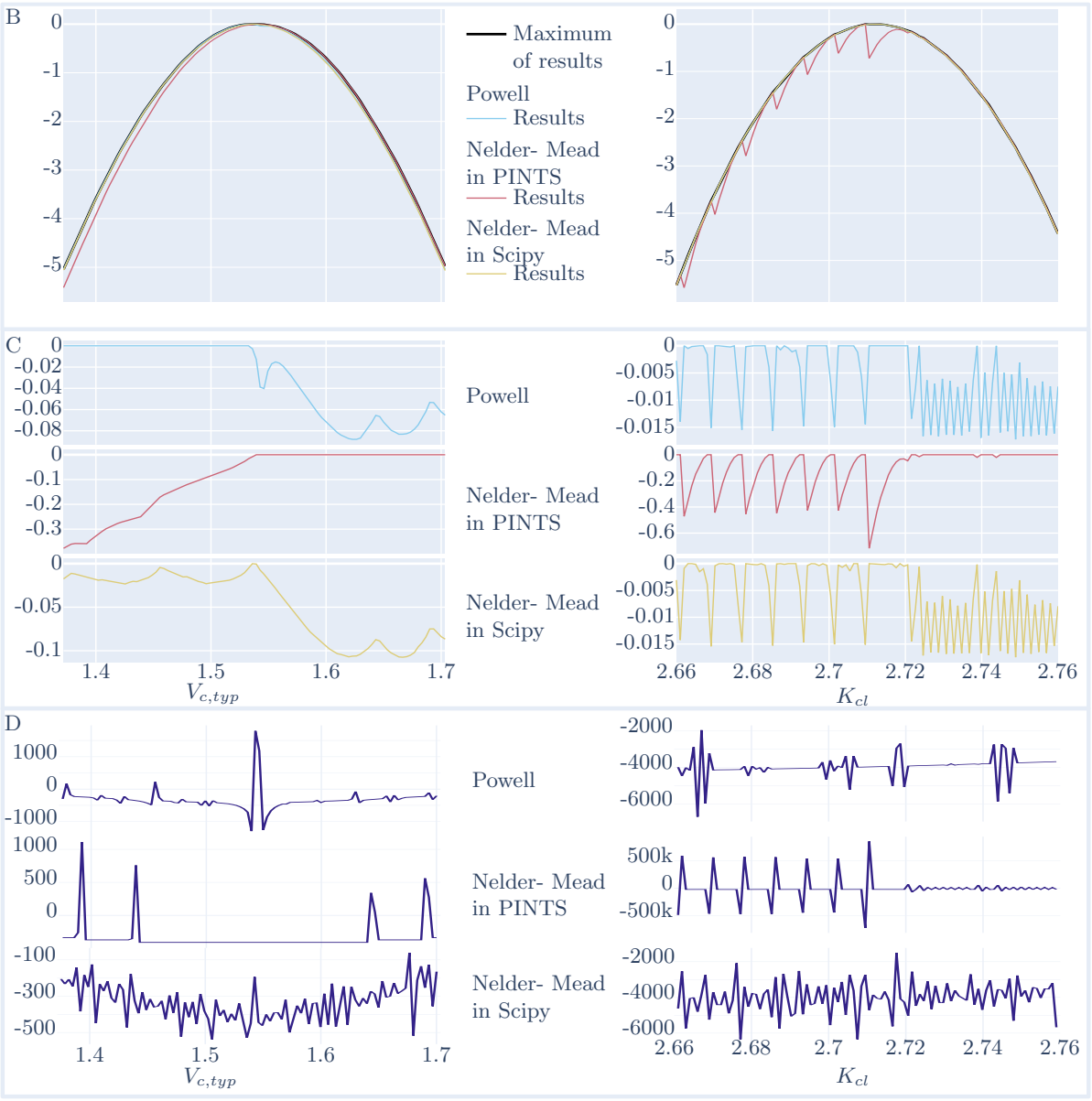



Figure 6.8: (A) Time to create profile likelihood graph and sum of squares error for different methods and different levels of extrapolation to determine next start point. (B) Profile likelihood from the three different methods with quadratic extrapolation. (C) The error for each method as the parameter of interest varies. (D) The second derivative of the results of each method. The true profile likelihood is assumed to be the maximum of all profile likelihoods and error is calculated as the difference from this true profile likelihood.

to $N_p$ 1-dimensional problems. If the optimum of each problem, $1 \leq i \leq N_p$, is $\theta^{(g)} + \sum_{\iota=1}^{i} \alpha_\iota s_\iota$ for some scalars $\alpha_\iota \in \mathbb{R}$, then the new point is selected as

$$\theta^{(g+1)} = \min_{\bar{\alpha}} \left( \theta^{(g)} + \bar{\alpha} \sum_{\iota=1}^{N_p} \alpha_\iota s_\iota \right),$$

for $\bar{\alpha} \in \mathbb{R}$. The search vectors are then updated such that the most successful vector (the vector that resulted in the largest step in objective function) is removed and $\sum_{\iota=1}^{N_p} \alpha_\iota s_\iota$ is appended to the list. [34]

Although Powell is a quick and reliable method, it takes 10 minutes to run for a single parameter in this simple mixed-effects model. However, it can be seen in Figure 6.6, when using CMAES, that the optimised parameters lie on a curved line. My current methodology starts at the MLE, then takes small fixed steps in $\theta_k$ towards the edges of the parameter space. At each step, $g$, the remaining parameters, $\theta_{-k}^{(g)}$, are optimised using $\theta_{-k}^{(g-1)}$ as a starting point. This could be considered a zeroth order extrapolation at each step. If the starting point was closer to the optimum value then less steps would be taken in optimisation, and a first order (linear) or second order (quadratic) extrapolation could lead to a closer starting point. I have tested both the Powell and Nelder-mead methods under these extrapolations, shown in Figure 6.8A. There is no improvement in the error of the results, and there is only slight improvement in the time taken for the analysis when using the Powell optimiser.

This analysis has lead me to develop an alternative method to find the profile likelihood, replacing the brute force method.

## 6.2 New methodology

Our aim is to determine practical identifiability using the profile likelihood method, in an efficient manner. It is not essential to have precise results but what is important information is the maximum of the curve, the confidence intervals and the shape of the curve.

We assume the maximum of the curve has already been found using maximum likelihood estimation. Next is to find the confidence intervals (C-I) of each parameter. This is equivalent to solving the equation

$$PLL_k(\beta) - LL\left(\hat{\theta}\right) + \frac{q_1(1-\alpha)}{2} = 0 \tag{6.4}$$

for $\beta$ to acquire the end points of the C-I of $\theta_k$. I propose a method based off of the method proposed by Venzon and Moolgavkar [60] but utilising the efficient local optimiser, Powell, as well as the quadratic approximation of the parameter space.

### 6.2.1 Algorithm

Initially, only one point of reference is known in the parameter space, $\hat{\theta}$, and thus we cannot approximate the shape of $PLL_k$. For the first step we assume that there is no correlation between $\theta_k$ and any other parameters, i.e. if $\theta_{-k}$ satisfies equation 6.1 for some $\theta_k = \beta$ then $\theta_{-k} = \hat{\theta}_{-k}$, the value of $\theta_{-k}$ at the MLE.

**First Step.** Solve

$$LL\left([\beta^{(0)}, \hat{\theta}_{-k}]\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2} = 0$$

for $\beta^{(0)}$. Let $\beta_L^{(0)}$ be the lower solution to this and $\beta_U^{(0)}$ be the upper solution. Numerically solving this problem is much easier as there is no need for optimisation or differentiation.

Now we optimise equation 6.1 to find $PLL_k\left(\beta^{(0)}\right)$ and $\theta_{-k}^{(0)}$. If the no correlation assumption is correct, then $\theta_{-k}^{(0)} = \hat{\theta}_{-k}$, $\beta^{(0)}$ satisfies equation 6.4 and no further iterations are required. If our assumption is incorrect, then we have acquired a point between $\hat{\theta}$ and the C-I end point where $PLL_k\left(\beta\right) - LL\left(\hat{\theta}\right) + \frac{q_1(1-\alpha)}{2} > 0$.

**Iterative Steps.** With three points, a quadratic approximation to the profile log-likelihood and the optimised parameter space can be made. We iteratively find the roots of the equation

$$\tilde{PLL}\left([\beta^{(g-1)}]\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2} = 0$$

where $\tilde{PLL}\left([\beta^{(g-1)}]\right)$ is the quadratic which goes through the points $\left(\beta_L^{(g-1)}, PLL_k\left(\beta_L^{(g-1)}\right)\right)$, $\left(\hat{\theta}_k, LL\left(\hat{\theta}\right)\right)$ and $\left(\beta_U^{(g-1)}, PLL_k\left(\beta_U^{(g-1)}\right)\right)$. At these approximate C-I points, optimise equation 6.1, starting at the point $\tilde{\theta}_{-k}^{(g-1)}$, the quadratic approximation of $\theta_{-k}^{(g)}$ from the points $\left(\beta_L^{(g-1)}, \theta_{-k,L}^{(g-1)}\right)$, $\hat{\theta}$ and $\left(\beta_U^{(g-1)}, \theta_{-k,U}^{(g-1)}\right)$. This gives $PLL_k\left(\beta^{(g)}\right)$ and $\theta_{-k}^{(g)}$.

**Termination.** There are three criteria for termination of this algorithm that can happen separately on the lower and upper C-I. The first is the Identifiable criterion, where the confidence interval was successfully found, i.e.

$$|PLL_k\left(\beta^{(f)}\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2}| \leq T_I$$

for the final iteration $f$ and some threshold $T_I$. The second criterion is the unidentifiable criterion, where the confidence interval has not been found and the Profile log-likelihood is stationary or increasing away from the confidence interval. This can

be tested as either a small or negative gradient when below the found MLE, or a small or positive gradient above the MLE, i.e.

$$T_\infty \geq \begin{cases} \frac{PLL_k\left(\beta_L^{(g-1)}\right) - PLL_k\left(\beta_L^{(g)}\right)}{\beta_L^{(g-1)} - \beta_L^{(g)}} & \text{for the lower C-I} \\ -\frac{PLL_k\left(\beta_U^{(g-1)}\right) - PLL_k\left(\beta_U^{(g)}\right)}{\beta_U^{(g-1)} - \beta_U^{(g)}} & \text{for the upper C-I} \end{cases}$$

for all $f - m_\infty < g \leq f$, some threshold, $T_\infty$, and some number of diverging iterations, $m_\infty$. This inequality ensures a strictly positive gradient below the MLE and a strictly negative gradient above. The final criterion is the maximum iteration criterion, where $f = m_{max}$. For this criterion there is not enough information to determine identifiability and the shape of the profile log-likelihood should be examined.

**Shape.** With the confidence intervals calculated, we can approximate the shape of the profile log-likelihood as a quadratic. However, as with the FIM, this is likely to not be accurate in every case, especially where there was unidentifiable termination. To determine the shape, $N_{shape}$ points, $\beta^{(f+s)}$ for $1 \leq s \leq N_{shape}$, are chosen. Equation 6.1 is optimised, starting at the points $\tilde{\theta}_{-k}^{(f+s)}$. A spline is then used to draw the profile likelihood through these points.

**Inaccuracies.** As the Powell optimisation method is a local optimiser, it can sometimes find vastly incorrect solutions. To solve this issue, during the iterative steps, if the optimised parameter has diverted far away from the expected parameter then a Global optimiser is used to confirm the optimised value. The iterations, $g$ where this occurs is when

$$\frac{1}{N_\theta - 1} \sum_{\kappa \neq k} \frac{\theta_\kappa^{(g)} - \tilde{\theta}_\kappa^{(g-1)}}{\tilde{\theta}_\kappa^{(g-1)}} < 3$$

where $N_\theta$ is the number of parameters of the model. The cutoff value of 3 was chosen as I had observed the above equation mostly lying between 0.5 and 2.

### 6.2.2 Testing Efficiency

I have tested this method on two model categories, logistic growth models and mixed-effects pharmacokinetic (PK) models.

The logistic growth models are of the form $\frac{\mathrm{d}C}{\mathrm{d}t} = rCf(C)$, where $C(t)$ is the population density at time, $t$, $r > 0$ is the growth rate and $f(C)$ is a capacity limiting function. This ODE has initial condition $C(0) = C_0$ These models were chosen as they are simple models but, depending on the form of $f(C)$ have known identifiability problems. There are three model forms that have been previously analysed by Simpson *et. al.* [55]:

- A standard logistic growth model, with $f(C) = \left(1 - \frac{C}{K}\right)$, for a carrying capacity $K > 0$.
- The Gompertz growth model, with $f(C) = \log\left(\frac{C}{K}\right)$.
- The Richards growth model, with $f(C) = \left(1 - \frac{C}{K}\right)^{\alpha}$ for some $\alpha > 0$

The previous paper uses Profile Likelihood methods to determine the identifiability of the parameters. The standard logistic growth model was found to be identifiable, however $C_0$ is unidentifiable in the Gompertz model and $K$ is the only identifiable parameter in the Richards' model. I have used the MLE parameters found in the Simpson *et. al.* paper [55] to generate a synthetic data-set for each of the models, and used two different methods to calculate the profile log-likelihoods from these data-sets: the new Quick Profile Log-Likelihood (QPLL) method, and the previous brute force method. This is to determine whether the results can be reproduced and whether there is any speed increase with the QPLL method.

When performing the two methods on the growth models, the time differences are not consistent, as can be seen in Fig 6.9. For all except one of the identifiable curves, QPLL was more efficient than sequential calculation. However, $C_0$ in the standard model was identifiable, and thus the algorithm was terminated before the maximum number of iterations were reached, but this did not improve the time taken to calculate. One thing that did occur was the global optimisation of one of the shape points which may have been a significant slowing factor. For the other Profile Log-likelihoods, where the identifiabilty could not be determined or was unidentifiable, there was mostly a decrease in efficiency. For $r$ and $C_0$ in the Richards' model, the algorithm didn't terminate. The curve is very steep below the MLE, as it approaches zero and is very flat above the MLE but still decreasing, thus not satisfying the unidentifiable termination criterion. In these cases the quadratic curve does not adequately describe the profile Log-likelihood. Points below zero would be suggested for the lower confidence interval which would have to be adjusted to above zero for the analysis, this then may have a value below the C-I line. As the quadratic approximation does not take into account the degree of steepness around this point, the next proposed point would be a position further away from the zero point. For the upper confidence interval the slope may be better described by a linear or exponential decay than a quadratic. The quadratic approximation fitted through the very steep lower confidence interval, causes much smaller proposed steps than appropriate for the likelihood. These are shown in Figure 6.10.

A

| Model | Parameter | Lower C-I | Upper C-I | Time Change |
|---|---|---|---|---|
| **Standard** | $r$ | Identifiable | Identifiable | -81.5% |
| | $K$ | Identifiable | Identifiable | -30.19% |
| | $C_0$ | Identifiable | Identifiable | 115.75% |
| **Gompertz** | $r$ | Identifiable | Identifiable | -62.54% |
| | $K$ | Identifiable | Identifiable | -18.66% |
| | $C_0$ | Unknown | Unknown | -35.86% |
| **Richards** | $r$ | Unknown | Unknown | 201.03% |
| | $K$ | Identifiable | Identifiable | -28.18% |
| | $C_0$ | Unknown | Unknown | 0.59% |
| | $\alpha$ | Unidentifiable | Identifiable | 52.72% |

B

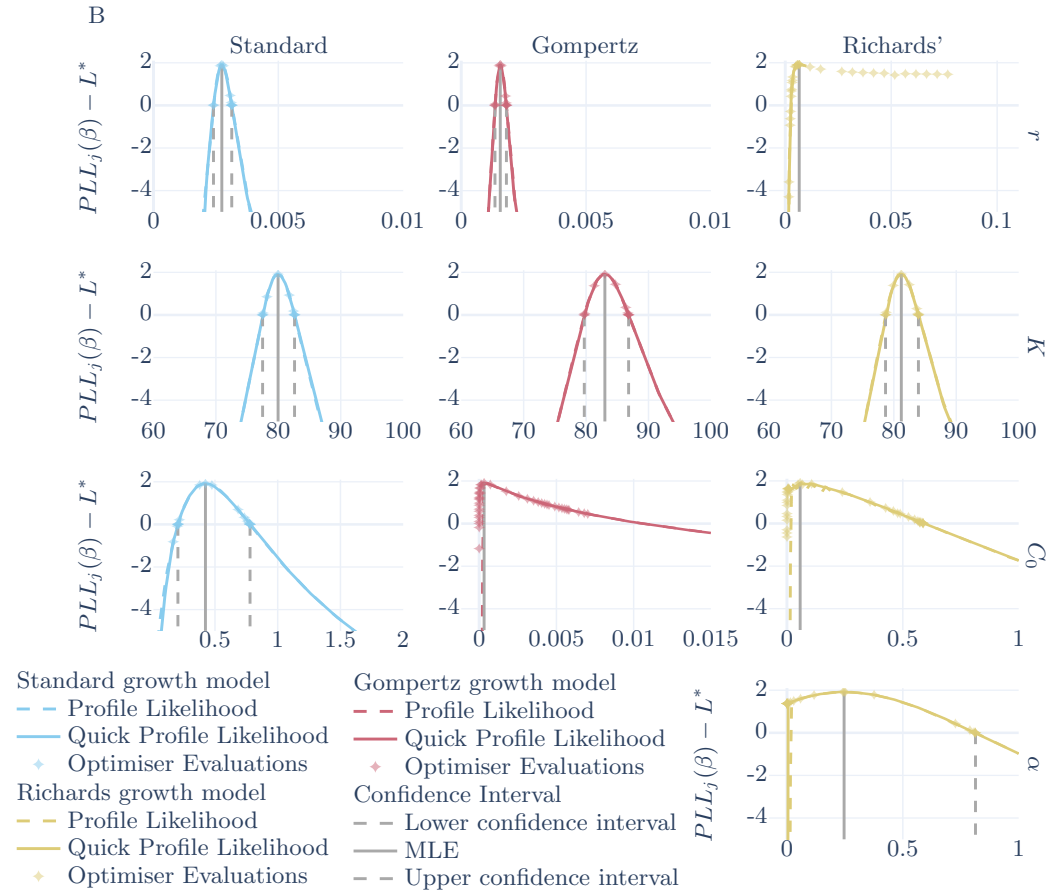



Figure 6.9: (A) Table showing the algorithm's determined identifiability and the time difference of the Quick Profile Log-likelihood method when compared with the sequential calculation of the Profile Log-Likelihood, for each parameter in the three models. The lower or upper confidence interval (C-I) is identifiable or unidentifiable if the algorithm was terminated by the identifiable or unidentifiable criterion resp., and is unknown otherwise. (B) Profile Log-likelihood curves, normalised by $L^* = LL\left(\hat{\theta}\right) - 1.92$. Dashed curve shows the result from the sequential calculation, solid curve is the result from QPLL and diamonds are the points the QPLL algorithm optimised at to find the confidence intervals. Where found, the QPLL confidence intervals are shown by a vertical dashed line and the MLE a solid vertical line.

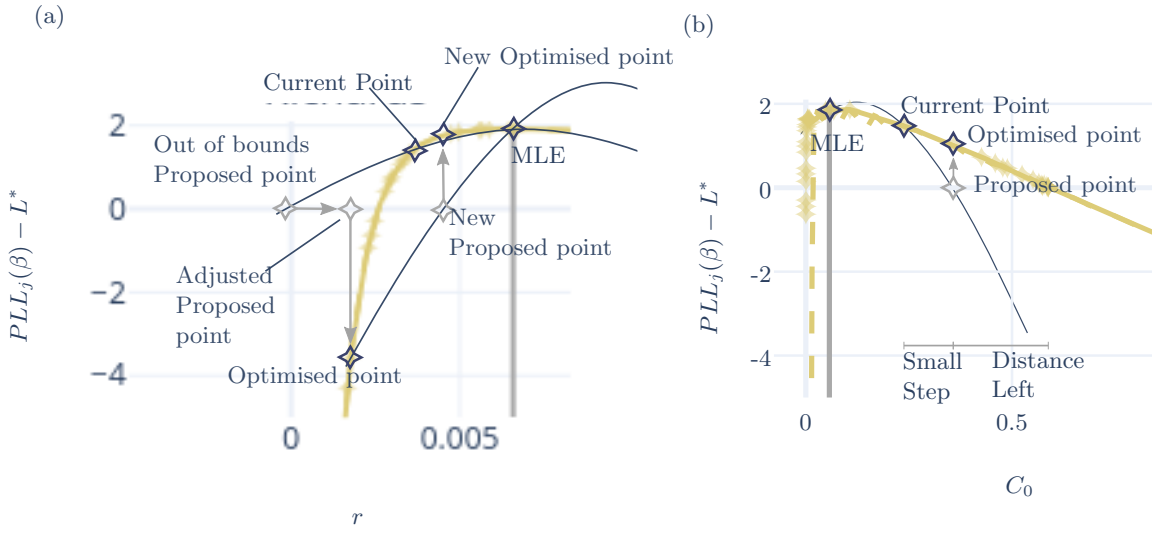



Figure 6.10: Proposed action of algorithm that leads to slow or non-convergence, without stationary or diverging termination being reached. Blue curves represent a potential fitted quadratic and the yellow curve is the profile log-likelihood. The graphs show the algorithm steps when finding (A) the lower confidence interval of $r$ , or (B) the upper confidence interval of $C_0$ in the Richards' model.

To combat this I implemented two updates to the algorithm. The first is a detection of this slow convergence. When several small steps have been taken, defined by

$$|PLL_k\left(\beta^{(g)}\right) - PLL_k\left(\beta^{(g-1)}\right)| \leq |PLL_k\left(\beta^{(g)}\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2}|,$$

for either lower or upper $\beta$ then the algorithm changes for the lower or upper half respectively. This new algorithm uses a second order extrapolation from the previous iterations, which is solved to find the crossing point of $LL\left(\hat{\theta}\right) - \frac{q_1(1-\alpha)}{2}$. If only the upper half has moved to this new method then the lower half ignores $\beta_U$ in its quadratic approximation and assumes the MLE is the maximum of the quadratic, and vice versa. The second update to the method simply ensures that if the algorithm returns a proposed point $\beta^{(g)}$ that is further from $\hat{\theta}_k$ than a previous iteration $\beta^{(\lambda)}$, $\lambda < g$ where $PLL_k\left(\beta^{(\lambda)}\right) - LL\left(\hat{\theta}\right) + \frac{q_1(1-\alpha)}{2} < 0$ then the proposed point is adjusted such that it lies between $\beta^{(\lambda_{max})}$ and $\beta^{(\lambda_{min})}$ where $\lambda_{max}, \lambda_{min} < g$ are the iterations at which

$$\beta^{(\lambda_{max})} = \max_{\lambda \in \Lambda_-}\left(\beta^{(\lambda)}\right),$$
$$\beta^{(\lambda_{min})} = \min_{\lambda \in \Lambda_+}\left(\beta^{(\lambda)}\right),$$

where,

$$\Lambda_- = \{\lambda | PLL_k\left(\beta^{(\lambda)}\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2} < 0\},$$

$$\text{and} \qquad \Lambda_+ = \{\lambda | PLL_k\left(\beta^{(\lambda)}\right) - LL\left(\hat{\theta}\right) + \frac{q_1\left(1-\alpha\right)}{2} > 0\}.$$

Implementing these changes did not improve the time efficiency but the adaptive algorithm did successfully find the C-I where they existed and returned unidentifiable otherwise.

The other model which these models have been tested on is a one-compartment linear PK model (Equation 2.1) with $V_C$ set as a mixed-effects parameter using data-set 5 from the category 3 data. The results from this is shown in Figure 6.11. This shows a significant improvement in efficiency with only a slight reduction in accuracy when shape points are included.

## 6.3 Conclusion

I have applied profile likelihoods to determine the practical identifiability of our models. However the brute force method to acquire the profile likelihood is incredibly inefficient, thus I explored techniques to improve this efficiency. Changing the optimiser to Powell from Nelder-Mead, and using extrapolation to determine the starting point for the optimiser, both had a slight improvement in the time taken to complete the analysis. However, for a greater improvement, I also developed a new method, the QPLL method, to acquire the profile likelihood. For models with a small number of parameters, where there may be a majority unidentifiable parameters, the QPLL method is significantly slower than the brute force method, which was already performing sufficiently fast. However, the adaptive QPLL can explore a large range of parameter values to find if the profile likelihood is unidentifiable or eventually crosses the confidence line. The brute force method only checks in the user defined window which may not be sufficient. For, highly complex models, such as mixed-effects models, where there are a large number of parameters to optimise, but the majority of parameters would be expected to be identifiable, because, for example, the simpler fixed-effects model is identifiable, then the QPLL method is the most efficient method. Using the profile likelihood, I have identified that the combined noise model and the multiplicative model are unidentifiable. Going foward I will not be using the combined noise model and fixing $\eta$ to one in the multiplicative model when analysing the experimental data. As explained in Chapter 4, this discovery is what lead to me using the constant noise model in generating the mixed-effects data-sets.

(a)

| | Method | Brute Force | QPLL | With shape points |
|---|---|---|---|---|
| TIME | $V_c$ | 14.3624 | 1.8226 | 2.961 |
| | $K_{cl}$ | 10.2902 | 0.9586 | 1.9063 |
| | **Total** | 24.6525 | 2.7811 | 4.8674 |
| ERROR | $V_c$ | 0.0 | 121.3334 | 0.2132 |
| | $K_{cl}$ | 0.0 | 0.0562 | 0.0368 |
| | **Total** | 0.0 | 121.3895 | 0.2499 |

(b)

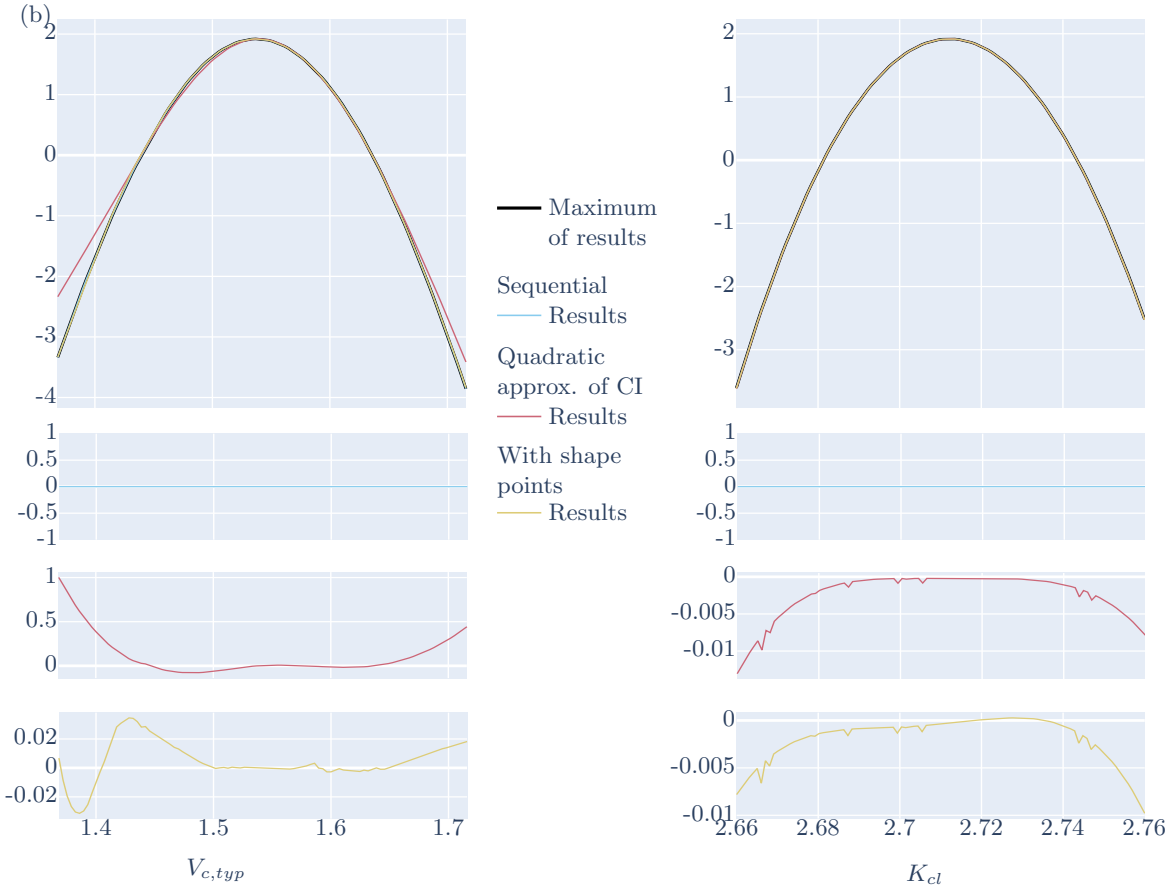



Figure 6.11: (A) Time to create profile likelihood graph and sum of squares error for the original sequential method and the quick Profile Log-Likelihood with and without 10 shape points (B) results from the different methods and the "error" when compared with the sequential results. (C) The second derivative of the results of each method.

# Chapter 7

# Bayesian inference

The identifiabilty of the models have been confirmed in Chapter 6 and in Chapter 5 the maximum likelihood estimate and maximum-a-posteriori estimate produced information on the most likely set of parameters, and thus the most likely drug response. However, as there is noise in the system, we cannot be sure this is the true parameter set, and redoing the experiment will lead to a different set of parameters. To provide full information on the uncertainty in these estimates, a Bayesian approach to parameter inference can be used. This approach finds the uncertainty by recovering the full posterior, $P\left(\theta|X^{\text{obs}}\right)$, instead of maximising it. In this section I will explain how Monte Carlo sampling can be used to recover the posterior and then perform this sampling on the fixed-effects and mixed-effects PKPD models.

To recover the posterior distribution, we return to the numerator of Bayes rule, $P\left(\theta|X^{\text{obs}}\right) \propto P\left(X^{\text{obs}}|\theta\right)P\left(\theta\right)$. The posterior probability distribution can then be estimated using Markov chain Monte Carlo (MCMC) sampling over $P\left(X^{\text{obs}}|\theta\right)P\left(\theta\right)$. This method is implemented by iteratively selecting a point, $\theta^{(g)}$, in the parameter space such that the collection of points $S = \{\theta^{(g)}|n^* \leq g \leq n_t\}$ approximates independent sampling from the posterior probability distribution. The selection of this new point depends on the particular MCMC sampling algorithm. An example of this is shown in Algorithm 1.

In this algorithm, an initial point, $\theta^{(0)}$, the covariance matrix, $\mathbf{C}$, and initial warm-up period, $n^*$, needs to be chosen. While performing Bayesian inference on PKPD models, I have chosen the initial point to be the maximum likelihood estimate. This point is already in the high probability region of the posterior so the algorithm should require a smaller initial warm-up period. The choice of covariance matrix, $\mathbf{C}$, can affect the effectiveness of the MCMC algorithm, where in most cases, the best results occur when $\mathbf{C}$ matches the covariance of the posterior distribution. When this posterior covariance is unknown, an adaptive covariance Markov chain (ACMC)

**Algorithm 1** Metropolis Random Walk MCMC

**for** $1 \leq g \leq n_t$ **do**
  $\hat{\theta} \sim \mathcal{N}\left(\theta^{(g-1)}, \mathbf{C}\right)$ ▷ Propose a potential point to include into our samples
  $r = \frac{P\left(X^{\text{obs}}|\hat{\theta}\right)P\left(\hat{\theta}\right)}{P\left(X^{\text{obs}}|\theta^{(g-1)}\right)P\left(\theta^{(g-1)}\right)}$ ▷ Compare the posteriors of $\hat{\theta}$ and the last sample
  $u \sim U\left(0, 1\right)$
  **if** $r > u$ **then** ▷ There is a probability of $\min\left(r, 1\right)$ that $\hat{\theta}$ is accepted as a sample
    $\theta^{(g)} = \hat{\theta}$
  **else**
    $\theta^{(g)} = \theta^{(g-1)}$
  **end if**
**end for**
**return** $S = \{\theta^{(g)}|n^* \leq g \leq n_t\}$ ▷ Return the collection of samples, ignoring the initial $i$ samples

can be used instead of the Metropolis Random walk. These algorithms differ as the covariance matrix $\mathbf{C}^{(g)}$ is calculated at each step from the previous samples acquired. The initial $n^*$ points from the sample are excluded as the MCMC algorithm is still travelling to the posterior distribution during this initial phase. To determine this initial phase, multiple MCMC chains are started at different points in parameter space, checking at each iteration whether the chains have converged into the same space. This is done through the $\hat{r}$ convergence measure [19]. The closer this measure is to one, the closer the chains are to convergence. The MCMC algorithm incorporating these is described in Algorithm 2.

I used the Haario-Bardenet MCMC algorithm as implemented in the Python package PINTS. The results on category 1 simulated PD data are shown in Figure 7.1. As can been seen, the posterior distribution for the PD model parameters approximately follows a Gaussian shape. However the noise parameters, $\sigma_{c,\text{PD}}$ and $\sigma_{m,\text{PD}}$, do not follow the same shape and have much wider probability distributions. This supports the results found in section 6, that these parameters are not identifiable.

For the experimental data, I performed Bayesian inference on the fixed-effects PKPD with either a constant noise model (Equation 2.23) or a multiplicative noise model (Equation 2.24) with $\eta$ fixed to one. The combined noise model was excluded and $\eta$ has been fixed as, from section 6, it was found that these noise models were unidentifiable. The results are shown in Figure 7.2. They both produced similar results and found a correlation between $MTT$, and $\gamma$. A slight correlation between $S$ and $MTT$ or $\gamma$ was found when using the constant noise model but this was not seen when multiplicative noise was assumed.

**Algorithm 2** Haario-Bardenet Adaptive covariance MCMC with stopping criterion

**for** $0 \leq c \leq n_{\text{chains}}$ **do**
$\theta_c^{(0)} = \theta_{\text{opt}} + \epsilon_c$ ▷ Choose initial points around the optimised parameters
$n_t = n_a$ ▷ The length of the non-adaptive stage before Adaptive covariance begins
$\mathbf{C} = \mathbf{C}^{(0)}$ ▷ Initial estimation of the Covariance
Do Algorithm 1 ▷ Perform non-adaptive covariance MCMC
**end for**
**while** $\hat{r} > 1.01$ **do** ▷ Start the adaptive phase
**for** $0 \leq c \leq n_{\text{chains}}$ **do**
$g = g + 1$
$\mathbf{C} = s_d Cov\left(\theta_c^{(0)}, \theta_c^{(1)}, \ldots, \theta_c^{(g-1)}\right) + s_d \epsilon_{cov} I_d$ ▷ Adapt the covariance for the proposed point
$\hat{\theta} \sim \mathcal{N}\left(\theta_c^{(g-1)}, \mathbf{C}\right)$ ▷ Propose a potential point to include into our samples
$r = \frac{P\left(X^{\text{obs}}|\hat{\theta}\right)P\left(\hat{\theta}\right)}{P\left(X^{\text{obs}}|\theta_c^{(g-1)}\right)P\left(\theta_c^{(g-1)}\right)}$ ▷ Compare the posteriors of $\hat{\theta}$ and the last sample
$u \sim U(0,1)$
**if** $r > u$ **then** ▷ There is a probability of $\min(r,1)$ that $\hat{\theta}$ is accepted as a sample
$\theta_c^{(g)} = \hat{\theta}$
**else**
$\theta_c^{(g)} = \theta_c^{(g-1)}$
**end if**
$X_{c,1} = \{\theta_c^{(\hat{g})} | g - n_s \leq \hat{g} \leq g - \frac{n_s}{2}\}$ ▷ Split the chain into half chains
$X_{c,2} = \{\theta_c^{(\hat{g})} | g - \frac{n_s}{2} \leq \hat{g} \leq g\}$ ▷ Where $n_s$ is the desired number of final samples
**end for**
$\hat{var} = \frac{n_s - 2}{n_s} W_s + \frac{2}{n_s} B_s$ ▷ Where $W_s$ is the mean of the variance within each $X_{c,j}$
$\hat{r} = \sqrt{\frac{\hat{var}}{W_s}}$ ▷ and $B_s$ is the mean of the variance between $X_{c,j}$
**end while**
**return** $S = \{\theta_c^{(g)} | n_t - n_s \leq g \leq n_t, 0 \leq c \leq n_{\text{chains}}\}$ ▷ Return the last $n_s$ number of samples

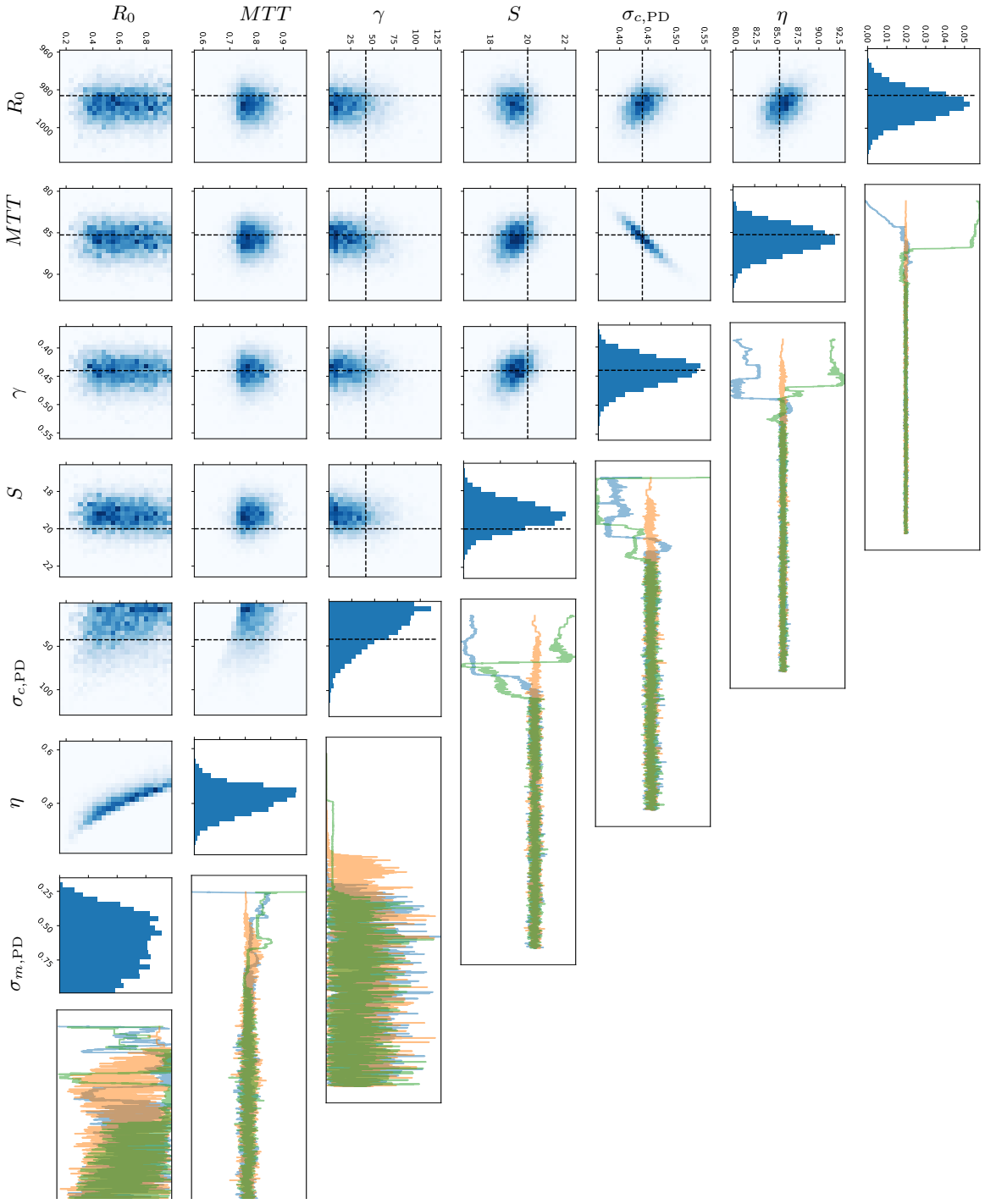



Figure 7.1: Results from the MCMC sampling as in Algorithm 2 on the PD model with simulated data when combined constant and multiplicative noise is assumed. The diagonal shows a histogram of the values for each parameter from the accepted samples. Below the diagonal are pairwise heat maps of parameter values from the samples. Next to these are traces of the different MCMC chains through the parameter space. Starting points for the chain are the Maximum likelihood estimate (MLE), the geometric mean of the MLE and upper bound, and the geometric mean of the MLE and lower bound.

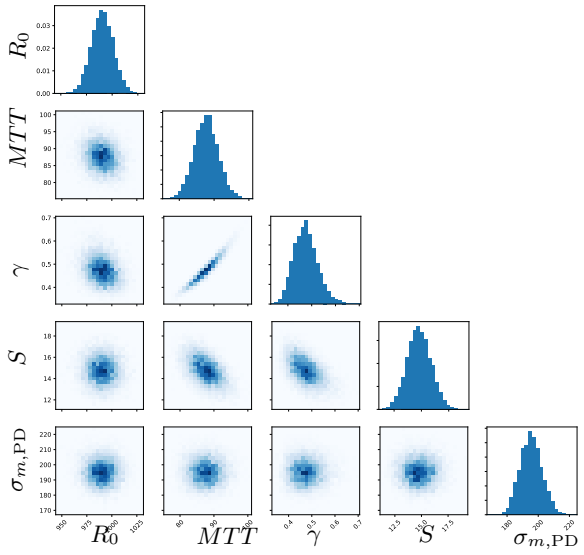
A

| param. | mean | std | $\hat{r}$ | ess |
|---|---|---|---|---|
| $R_0$ | 990.5 | 10.9 | 1.0013 | 2183.3 |
| $MTT$ | 87.8 | 3.5 | 1.0012 | 2363.5 |
| $\gamma$ | 0.48 | 0.05 | 1.0019 | 2299.2 |
| $S$ | 14.8 | 1.1 | 1.0004 | 3180.3 |
| $\sigma_{c,\mathrm{PD}}$ | 195.4615 | 7.6 | 1.006 | 950.2 |
| $\sigma_{m,\mathrm{PD}}$ | . | . | . | . |

B

| param. | mean | std | $\hat{r}$ | ess |
|---|---|---|---|---|
| $R_0$ | 987.8 | 11.1 | 1.002 | 2670.9 |
| $MTT$ | 90.6 | 3.7 | 1.001 | 2729.2 |
| $\gamma$ | 0.52 | 0.05 | 1.001 | 2697.9 |
| $S$ | 12 | 0.9 | 1.0 | 2397.5 |
| $\sigma_{c,\mathrm{PD}}$ | . | . | . | . |
| $\sigma_{m,\mathrm{PD}}$ | 0.2 | 0.0081 | 1.0004 | 2369 |

C




D

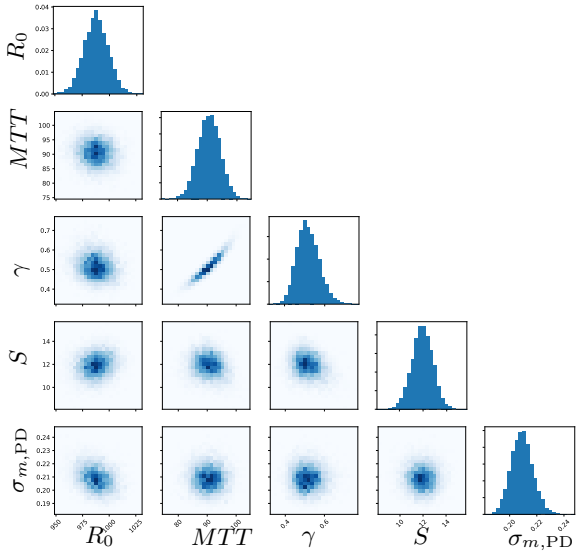



Figure 7.2: Results from the MCMC sampling using Algorithm 2 on the PD model with experimental data when either constant noise (A & C) or multiplicative noise (B & D) is assumed. A & B) Table of results with the mean value, the standard deviation, final value of $\hat{r}$ and effective sample size for each parameter from the accepted samples. C & D) The diagonal shows a histogram of the values for each parameter from the accepted samples. Below the diagonal are pairwise heat maps of parameter values from the samples.

## 7.1 Mixed-Effects

In the mixed-effects case, the posterior as defined in Chapter 5.3 can also be sampled from using Monte Carlo algorithms. In this case, I will use the NUTS (No U-Turn sampling) algorithm to sample from the posterior of the hyper-parameters and individual-level parameters simultaneously. The NUTS is a form of Hamiltonian Monte Carlo algorithm where instead of choosing a proposal point given a probability distribution, it simulates the sampler as a particle travelling with velocity upon the hyper-surface of the negative log-likelihood. This particle is then subject to Hamiltonian mechanics and proposal points are taken along it's path of travel. This particle naturally moves towards areas of high log-likelihood, meaning the acceptance rate of proposed points is high and convergence to the area of interest is quick [25]. This is incredibly useful for systems with a large number parameters and heavy correlation, such as mixed-effects models. However, each iteration of the NUTS algorithm takes much longer than Haario-Bardenet MCMC and not as many samples are able to be acquired in the same amount of time. A preliminary run of both algorithms on the PD mixed-effects model, using category 2 data, found that Haario-Bardenet MCMC did not converge, and had $\hat{r}$ values above 10 where NUTS had $\hat{r}$ values mostly between 2 and 4 for each parameter, in approximately 10% of the number of iterations. However this run with NUTS still took longer than the Haario-Bardenet run.

These $\hat{r}$ values show that, although better than Haario-Bardenet, NUTS was still under-performing. From the maximum-a-posteriori (MAP) (see Figure 5.3), values were inferred close to the data-generating parameters. However the NUTS chains do not sufficiently explore the parameter space. In order to solve this issue I altered the methods in two ways. The first was to use the runs of MAP, that had converged to the maximum, as the starting points of the different chains. As our aim in sampling is to examine the variability in our parameter estimates, this alteration ensures our chains start in, and sample from, the region of interest of the parameter space. As CMAES is also an efficient method for finding the optimum, it is likely to speed up the process of converging to the region of interest. Secondly, I decided to reduce the parameter space by fixing some of the PK parameters to the values found from running MAP on the two-compartment PK model only. I started with fixing $\sigma_{PK}$ then fixed $V_1$, $K_1$ and $K_{Cl}$, one after the other, in that order. The number of CMAES runs, out of ten, that converged to the optimum (a run with posterior value less than 1% away from the maximum posterior value found) are

- seven for the full set of parameters,

- six with $\sigma_{PK}$ fixed,
- eight with $\sigma_{PK}$ and $V_1$ fixed,
- nine with $\sigma_{PK}$, $V_1$ and $K_1$ fixed,
- nine with $\sigma_{PK}$, $V_1$, $K_1$ and $K_{Cl}$ fixed.

To perform the posterior sampling, I then chose to keep $\sigma_{PK}$ and $V_1$ fixed and use the eight converged MAPs as starting points for the sampler. These chains still had problems mixing. When additionally fixing $K_1$ and using the nine MAPs that were acquired in that respective estimation, the sampling chains did converge.

While keeping $K_1$, $\sigma_{PK}$ and $V_1$ fixed, the number of mixed-effects parameters can be increased. I have chosen $R_0$ to be the next parameter to introduce as mixed-effects, thus $\psi = [V_c, R_0]$. Running the maximum-a-posteriori on this model with the Category 2 data, with the same method used in the $\psi = [V_c]$ case, the 10 runs did not end with a log-posterior score of within 1.92 of each other. However, 8 of these runs still produced reasonable looking fits. Expanding the requirement of convergence to be within $\frac{q_{n_p}(0.95)}{2} = 62.73$ of the maximum, where $q_{n_p}(1-\alpha)$ is the $(1-\alpha)$th quantile of the $\chi^2$ distribution with $n_p$ degrees of freedom and $n_p = 101$ is the total number of parameters of the system, allows these 8 to be considered converged.

I then ran 5 chains of the NUTS sampler on the log-posterior of this model, as shown in Figure 7.4. The chains converged well for $R_{0,\text{typ}}$, $\omega_{R_0}$ and $R_{0,i}$, for individual $0 < i \leq n_i$, and correctly predicted small $\omega_{R_0}$, but with the prior stopping the value becoming zero. However, the sampler converged poorly for $V_{c,\text{typ}}$, $\omega_{V_c}$ and $V_{c,i}$ and there is some but not complete convergence in the fixed-effects parameters. This shows the importance of starting with a simpler model and incrementally adding complexity.

## 7.2 Conclusion

Using Monte Carlo Markov chains or the Hamiltonian-based sampler, NUTS, to sample from the posterior distribution, has produced a sample of probable parameter sets for each model, which, as well as being able to view the probability distributions of these parameters, will lead into making model choices in Chapter 8, and making predictions in Chapter 9. However, for the mixed-effects model there have been some convergence issues in the chains which required a simplification of parameter space and an adaption of the method to solve. This method of parameter inference

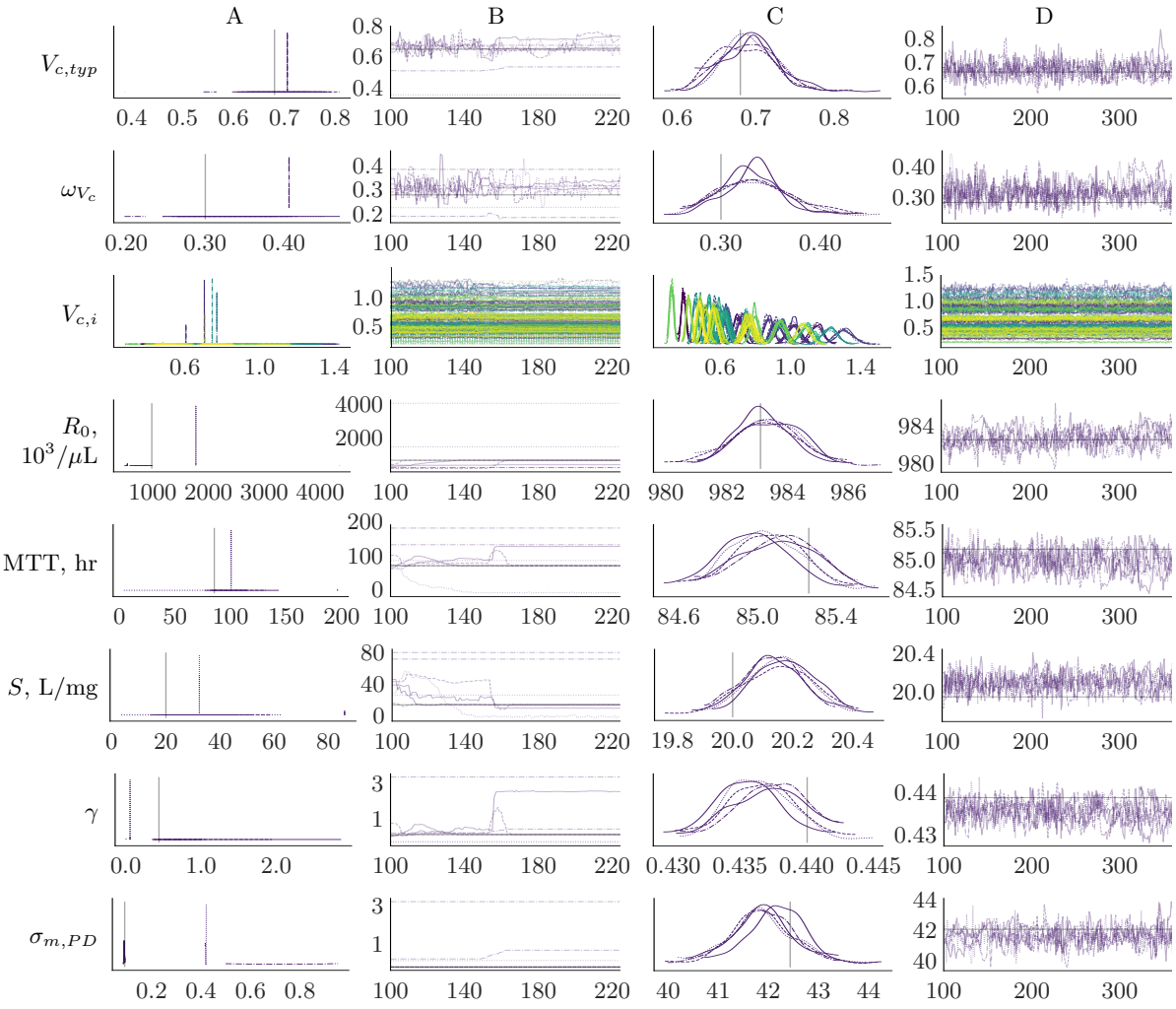



Figure 7.3: Sampling results comparing between no fixed parameters and random start (A & B) and, starting at the MPEs while fixing $\sigma_{PK}$, $V_1$ and $K_1$ (C & D). Shows the distribution of samples (A & C) and the trajectory of the samplers (B & D) for each parameter. Each line represents a different sampler and each colour in the $V_{c,i}$ plots represents an individual.

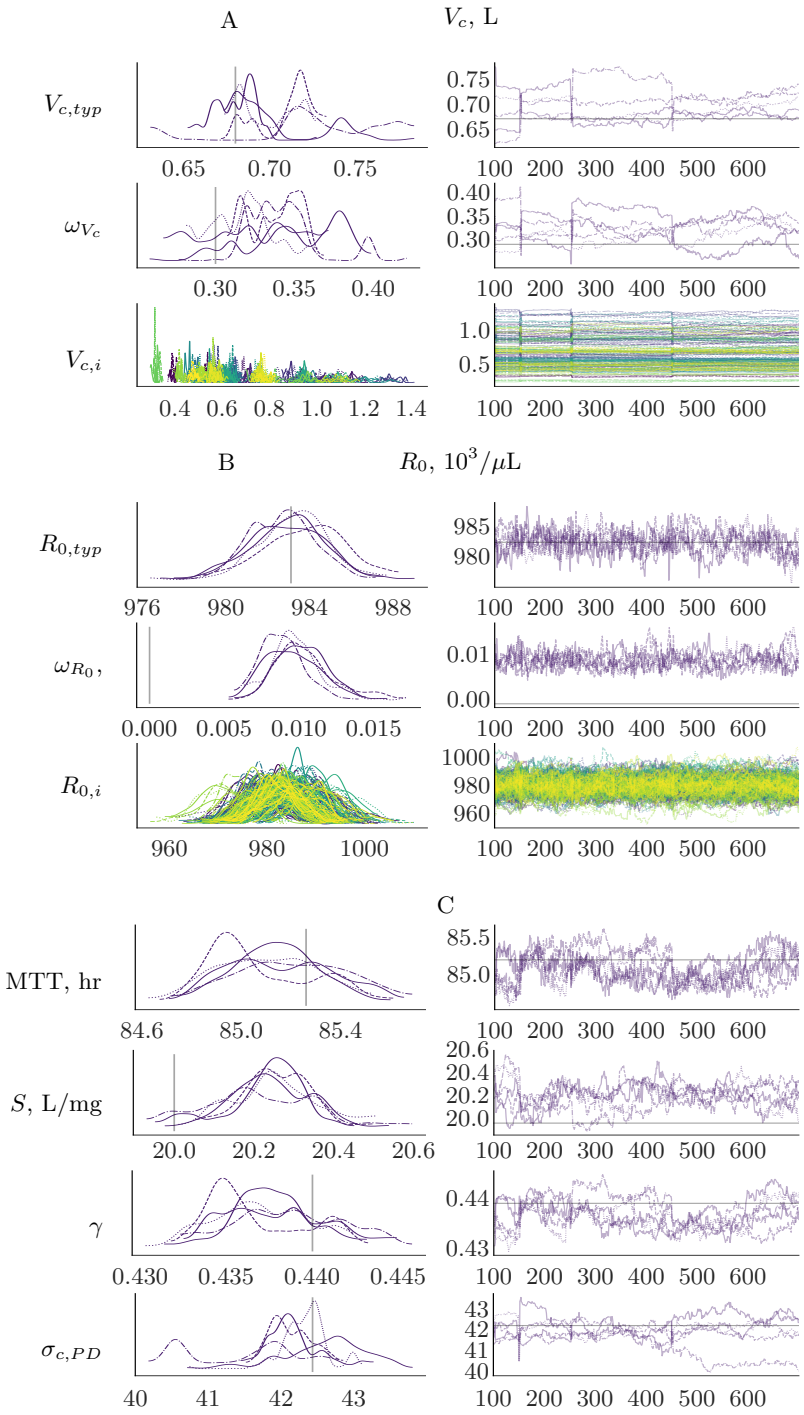



Figure 7.4: Sampling results when two parameters, $V_c$ (A) and $R_0$ (B), are modelled as mixed-effects, and the rest (C) as fixed-effects. Shows the distribution of samples (left) and the trajectory of the samplers (right) for each parameter. Each line represents a different sampler and each colour in the $V_{c,i}$ and $R_{0,i}$ plots represents an individual.

can take longer than the maximum likelihood estimation seen in Chapter 5 and the QPLL method from Chapter 6 but provides useful information about the parameter space which is utilised in further analysis.

# Chapter 8

# Model selection

The final step to the inverse problem is to select the most probable model that can best describe the drug response. We have refined each model and inferred parameters that optimised the probability that this model describes the data in Chapters 5 and 7. However this is under the assumption that the chosen model can sufficiently represent the system and is the best model at predicting different scenarios. In this section I will outline different ways to compare models, and explain how to quantify out-of-sample predictive power. I will then use this predictive power to assess which of the noise models is best and to build a mixed-effects population model.

When making model choices, it is important that the model is improved as these models become more complicated and include greater numbers of parameters. Commonly, a likelihood ratio test is used to determine between two nested models that differ in a single parameter. This was the method that Friberg et. al. used to make choices in the myelotoxicity model.

The likelihood ratio is defined as

$$\lambda_{LR} = -2\left(\sup_{\theta_0 \in \Theta_0}\left(LL\left(\theta_0\right)\right) - \sup_{\theta \in \Theta}\left(LL\left(\theta\right)\right)\right),$$

where $\Theta$ is the parameter space of the proposed model (the mixed-effects model in this case) and $\Theta_0$ is the parameter space of the null hypothesis (e.g. $\omega_{V_c} = 0$ when comparing a mixed-effects model to a fixed-effects model, or $\sigma_{m,\mathrm{PD}} = 0$ when comparing a constant noise model to a combined model).

The likelihood ratio test is based on Wilks' theorem which states that when the null hypothesis, $H_0$, is true then

$$T = -2\log\frac{L\left(\mu_0, \widehat{\theta}_0\right)}{L\left(\widehat{\mu}, \widehat{\theta}\right)}$$

converges to a $\chi^2$ distribution. This is only true under certain conditions [3].

Conditions required by the Likelihood ratio test are

- Asymptotic - Sufficient data are collected. However, PKPD trials often have quite sparse data.
- Interior - Only values of the parameters of interest, e.g. $[\omega_{V_c}]$, and nuisance parameters, e.g. $[V_{c,typ}, \theta_{-V_c}]$, that are not on the boundaries of their parameter space are admitted. However, in this case, our null model $\omega_{V_c} = 0$ and $\sigma_{m,\mathrm{PD}} = 0$, are at the boundary.
- Identifiable - Different values of the parameters specify distinct models. The structural identifiabilty has been confirmed in this case.
- Nested - The null hypothesis H0 is a limiting case of the general case hypothesis H1, for example, with some parameter constrained to a subrange of the entire parameter space. This is true when comparing mixed-effects models where there is a single parameter changed between mixed and fixed-effects or when reducing a noise model. However, this is not necessarily true between any population or noise model.
- Correct - The true model is specified either under H0 or under H1. This is unknown.

For the above reasons it would not be best practice to use the likelihood ratio test for this. It would be better to use an alternative test. Preferably, this test will be an indicator of whether the model can predict circumstances outside of the data provided, such as the drug-response in a patient that is not in our sample. Gathering more real-life data to test model predictions is expensive and time-consuming. Thus other methods that can estimate out-of-sample predictive power without extra data are valuable. Leave-One-Out (LOO) cross validation is one. For this method, a single point, $x_j \in X^{\mathrm{obs}}$, is excluded, then the model is fitted to each set $X^{\mathrm{obs}}_{-j} = \left\{x^{\mathrm{obs}}_\zeta \in X^{\mathrm{obs}} | \zeta \neq j\right\}$, using MCMC sampling to provide the posterior distribution, $P\left(\theta | X^{\mathrm{obs}}_{-j}\right)$. The closeness of the excluded point, $x_j$, to the fitted curve is assessed, providing the probability, $P\left(x_j | X_{-j}\right)$. This is repeated for each $x_j \in X^{\mathrm{obs}}$, and the expected log point-wise predictive density (elpd)[58] is calculated as

$$\mathrm{elpd}_{\mathrm{loo}} = \sum_{j=1}^{n_{\mathrm{obs}}} \log P\left(x_j | X_{-j}\right). \tag{8.1}$$

The model with the highest elpd score is the model with the greatest predictive power. This method can be very computationally taxing and may not be feasible to perform. There are other methods that can estimate this score and are quicker to compute. $K$-fold cross validation is similar to LOO cross validation, however a set $X_k^{\text{obs}} \subset X^{\text{obs}}$ of size $K$ is excluded instead of a point. Similarly, the model is fitted to each set $X_{-k}^{\text{obs}} = \left\{x_\zeta^{\text{obs}} \in X^{\text{obs}} | x_\zeta^{\text{obs}} \notin X_k^{\text{obs}}\right\}$, and the elpd is calculated as

$$\text{elpd}_{\text{k-fold}} = \sum_k \sum_{x \in X_k^{\text{obs}}} \log \frac{1}{|S^{-k}|} \sum_{\theta^s \in S^{-k}} P\left(x|\theta^s\right), \tag{8.2}$$

where $S^{-k}$ is the set of parameters, $\{\theta_c^{(g)} | n_t - n_s \leq g \leq n_t, 0 \leq c \leq n_{\text{chains}}\}$, acquired from performing MCMC sampling on the subset of data $X_{-k}^{\text{obs}}$ [58]. Less fitting is needed for this method and thus it is quicker to compute. LOO could also be approximated from a single MCMC run by using importance sampling, with

$$P\left(x_j|X_{-j}\right) \approx |S| \left(\sum_{\theta^s \in S} \frac{1}{P\left(x_j|\theta^s\right)}\right)^{-1},$$

where $S$ is the set of parameters, $\{\theta_c^{(g)} | n_t - n_s \leq g \leq n_t, 0 \leq c \leq n_{\text{chains}}\}$, acquired from the MCMC sampling performed in Chapter 7. However the tails of the posterior from the parameter inference, $P\left(\theta|X^{\text{obs}}\right)$, are mismatched with the tails from the LOO posterior, $P\left(\theta|X_{-j}^{\text{obs}}\right)$, due to a difference in data size. Unfortunately, this can induce instability. This can be mitigated by using Pareto smoothed importance sampling (PSIS-LOO) which adjusts the shape of the tail to fit more with the LOO posterior [59]. Alternatively, the elpd can be approximated by the Widely Applicable Information Criteria (WAIC) which balances predictive accuracy and effective number of parameters,

$$\text{elpd}_{\text{waic}} = \sum_{x \in X^{\text{obs}}} \log \left(\frac{1}{|S|} \sum_{\theta^s \in S} P\left(x|\theta^s\right)\right) - p_{\text{waic}},$$

where $p_{\text{waic}}$ is the estimated effective number of parameters [62]. Similarly to PSIS-LOO, the only inputs required are the samples, $S$, from MCMC sampling during parameter inference and no further MCMC sampling is required.

Both PSIS-LOO and WAIC can encounter problems. If the posterior variance of the log predictive densities in WAIC is too large (commonly larger than 0.4) then WAIC becomes unreliable. In PSIS-LOO, if the estimated shape parameter $k$ of the generalised Pareto distribution is greater than 0.7 then the variance of the PSIS estimate is finite but may be large, leading to poor performance. To check the results

of one of these methods, the other method can be used. However if both methods encounter these problems, then it would be beneficial to check the results using K-fold cross validation.

## 8.1 Noise model

The samples acquired from the MCMC sampling were analysed using the Arviz package to determine the PSIS-LOO and WAIC elpd estimations on the fixed-effects Friberg model (Equations 2.17-2.21) when using category 1 simulated data, and compared the elpd for different noise models (Equations 2.23-2.25). The multiplicative Gaussian noise model had the greatest elpd score at -3131.3. This was not significantly different from the combined noise model, which had an elpd score of -3131.7. Both the PSIS-LOO and WAIC estimations found the same elpd score. The constant noise model had a significantly worse PSIS-LOO elpd of -14196.5 (and WAIC elpd of -16677.8). Thus it can be concluded that for the simulated data case either a multiplicative or combined noise model would be best used for future predictions. As it was shown in Chapter 6 that some parameters were unidentifiable when combined noise was assumed, it would be best to use the multiplicative noise model.

In the experimental data, the WAIC elpd for constant noise was higher by 11.909 (at a p-value of 0.08) compared to multiplicative noise. However, in the calculation of WAIC the posterior variance of the log predictive densities exceeded 0.4 which could be an indication of the WAIC estimation deviating from the true elpd. These results are supported by the LOO-PSIS estimation which had similar elpd values and standard error for constant and multiplicative noise. Thus in comparing the noise models, constant noise seems the most appropriate for our data, with fully identifiable parameters and a better WAIC score than multiplicative noise although only to a p-value of 0.08. This analysis is why constant noise was chosen for generating the mixed-effects models and used for analysis in the mixed-effects models.

## 8.2 Population model

With the samples acquired from the MCMC sampling I again used the Arviz package to determine the PSIS-LOO and WAIC elpd estimations on the Friberg model (Equations 2.17-2.21) with constant noise, (Equation 2.23) when using category 2 simulated data, and compared the elpd for three different mixed-effects models, $\psi \in \{[], [V_c], [V_c, R_0]\}$. In the case that matches the data, $\psi = [V_c]$, the elpd score was

significantly better than the other cases. The fixed-effects problem, $\psi = []$, performed the worst, with ELPD difference of $\Delta\text{elpd}_{\text{loo}} = 1690.6$ compared with the $\psi = [V_c]$ case and standard error $SE_\delta = 150.4$, showing that the mixed-effects model can be recovered. The model with extra mixed-effects parameters, $\psi = [V_c, R_0]$, did worse than the $\psi = [V_c]$ case with $\Delta\text{elpd}_{\text{loo}} = 153.5$, $SE_\delta = 5.89$ but was significantly better than the fixed-effects case. This shows that, in this case, the removal of required parameters has a greater effect on the score than the addition of unnecessary parameters. The WAIC metric also confirms these results, however in both cases the calculation encountered potential unreliability errors.

## 8.3 Conclusion

In conclusion, likelihood scores is not a suitable comparative measure for these models as it can lead to over-fitting. However, using the posterior samples acquired from Chapter 7, more meaningful comparative tools can be used to determine the most suitable model. The LOO-PSIS score estimates leave-one-out testing from the posterior samples without requiring multiple runs of the analysis on the data-sets with missing data. The WAIC score analyses the parameter samples and balances the fit to data while penalising the number of parameters of the model. From running these analyses I have found that the constant noise model outperforms the multiplicative noise in the experimental data. When analysing the population model on synthetic data I showed the true model is the most appropriate model but there were some warnings raised by the algorithm which indicate potential inaccuracies in the scores.

# Chapter 9

# Predictions

Now that the inverse problem has been addressed, we can return to the forward problem and simulate drug response under different scenarios. In this section, I will run simulations of the PKPD model in the experimental conditions, in order to compare with the observed data. Then, I will simulate the drug response under different dose amounts in order to determine at what dosage would the drug be considered toxic.

To determine whether a final model that has been settled on after doing all the above analysis, can adequately predict the data, the goodness of fit of this model can be assessed from the plot in Figure 9.1A. Plotting the observed platelet count versus the predicted platelet count using the maximum likelihood estimated parameters, shows the extent to which simulated predictions deviate from the data. Ideally, this would be equally distributed above and below the identity, predicted = observed. Plotting the local regression line using Locally WEighted Scatterplot Smoothing (LOWESS) shows that for my case, the fixed-effects model with constant noise is fitting well close to the baseline platelet count but under-predicts as it deviates away from this value. When the correct model is used, as in Figure 9.1C, where the category 2 simulated data is compared with simulations from the generating model with MAP parameter estimates, the regression line fits very closely with the identity.

The overall aim of PKPD modelling is to use it to support clinical decisions. PKPD models along with the inferred parameters can be used to determine the nadir (the lowest blood cell count) for each dose which is seen in Figure 9.1B. Thrombocytopeania is categorised into different grades depending on severity. Grade 1 is a platelet count in the range of $(150,000/\mu L, 75,000/\mu L]$; Grade 2 is a platelet count in the range of $(75000/\mu L, 50000/\mu L]$; Grade 3 is a platelet count in the range of $(50000/\mu L, 25000/\mu L]$; Grade 4 is a platelet count $< 25000/\mu L$. Using our maximum likelihood estimate leads to a prediction that Grade 1 thrombocytopeania will occur

at 66.9 mg/Kg of Docetaxel. This value seems unusually high, probably due to the different physiology of rats and humans. Kesteren *et. al.*, who performed a similar analysis with data from humans [57], found a baseline platelet count of 282,000/$\mu$L, much lower than found in this study. Using the ratio between these numbers to scale the ranges, a prediction of thrombocytopeania at dose 19.1 mg/Kg can be made. However, this does not incorporate the uncertainty we have in our parameter estimates. Simulating from 10% of the MCMC samples, leads us to determine that 11.5 mg/Kg is a safer dose and would avoid thrombocytopeania for those sample parameters. With more data, such as the category 2 simulated data, we can be more certain in our results and as such determine that 13.45 mg/Kg is still safe, as shown in Figure 9.1D.

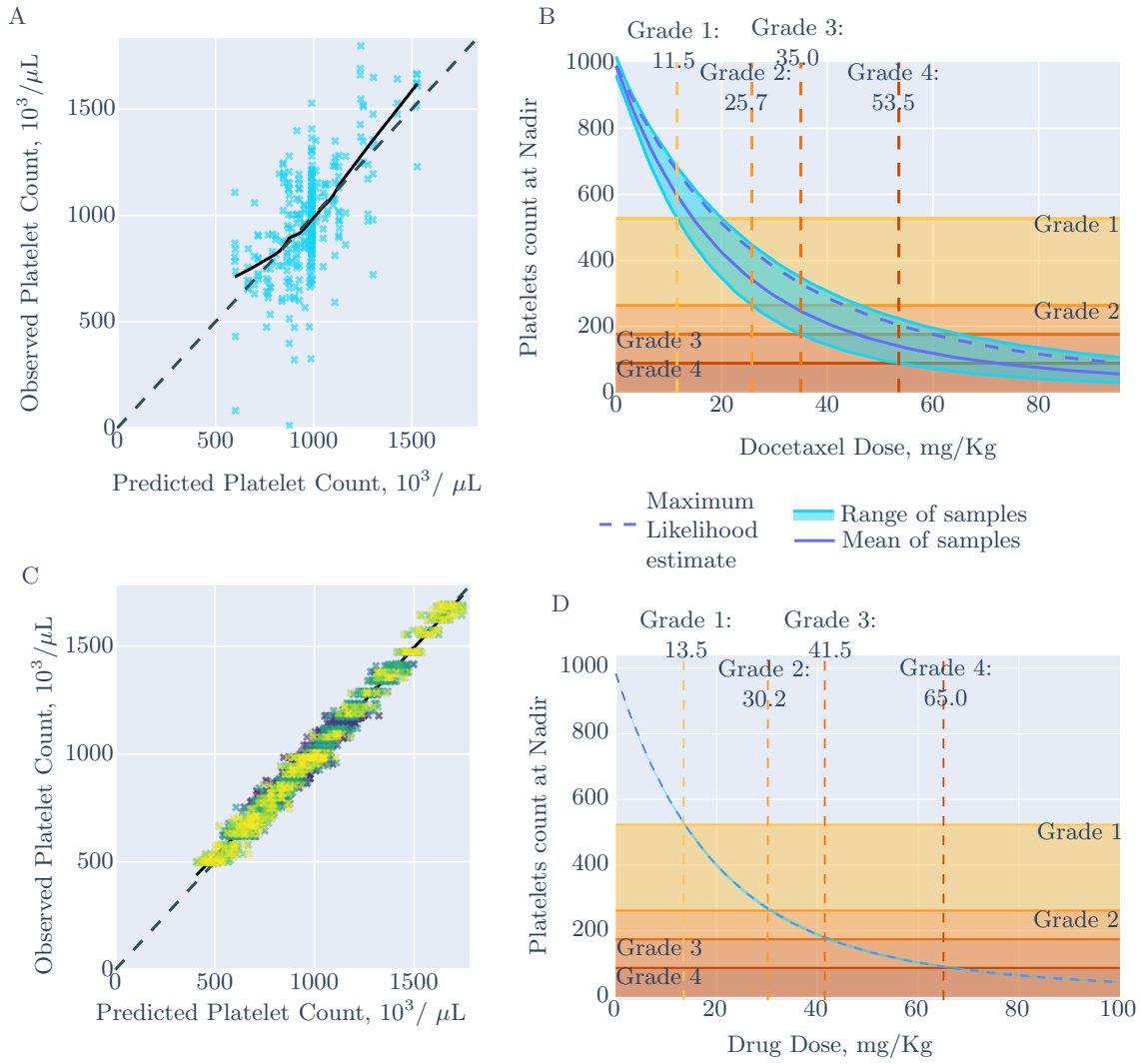



Figure 9.1: A & C) Goodness of fit graph comparing the predicted platelet counts from the maximum likelihood estimate to the observed platelet counts. Black line shows the local regression line using LOWESS regression with smoothing parameter of 0.6666 (the fraction of the total number of data points that are used in each local fit). Dashed line is the identity. B & D) Prediction of occurrence of thrombocytopeania. Shows the simulated platelet count at the Nadir when different levels of dose are administered and compares the simulation from 3674 MCMC parameter samples (10% of the fixed-effects samples), the mean of the parameter samples and the maximum likelihood estimation. Includes the point at which the minimum nadir of the samples would be considered Grades 1 to 4 thrombocytopeania. A & B) Predictions of the fixed-effects PKPD model, (Equations 2.2-2.3 and 2.17-2.21). with constant noise, (Equation 2.23). when analysis is performed on the experimental data. C & D) Predictions of the same model with mixed-effects parameters, $\psi = [V_c]$, when analysis is performed on category 2 simulated data.

# Chapter 10

# Conclusion and Discussion

My goal with this thesis is to outline a standard procedure for PKPD analysis and to test various methods for this purpose. Maximum likelihood estimation or maximum-a-posteriori estimation is an easy way to infer the parameters of a system and was the fastest to compute. However, it does not provide important information such as the uncertainty in these parameter estimates, or the degree of parameter identifiability, which can affect our conclusion. It can give an indication of goodness of fit for model comparison using the maximum likelihood scores of each model. However, as discussed in section 8 these methods are not appropriate for most PKPD analyses. This estimate does serve as a good starting point for both acquiring the profile likelihoods and MCMC sampling, to speed up those methods. It can also give an initial indication of potential difficulties in optimisation and sampling if the fitted curve or parameter values are not as expected. In our case the fact that $\sigma_{m,\mathrm{PD}}$ is at the upper bound for multiplicative noise and that combined noise reduced $\sigma_{m,\mathrm{PD}}$ to almost zero indicated that $\sigma_{m,\mathrm{PD}}$ was potentially not appropriate for this data.

The use of profile likelihoods to explore parameter identifiability provides a way to view the parameter space and give an indication on how unidentifiable parameters could be resolved. It also provides a view on the relation between parameters. However, the disadvantage to this method is the time it takes to compute. With the new algorithm for determining profile likelihoods, this time is significantly reduced for larger complicated systems, allowing it to be used more regularly. However, further development of this method can still be done to increase efficiency, especially in the case of non-identifiable parameters and unusually shaped profile likelihood curves.

Using Bayesian inference and MCMC sampling, the uncertainty in parameter estimates are identified. This allows the variability of parameters to be included in any predictions made from the models. Correlated parameters can also be seen in the pairwise plots of samples. This method does take more time than the maximum

likelihood estimate but for the fixed-effects model completed in a reasonable amount of time. Convergence with mixed-effects parameters takes much longer or does not occur at all, due to the increased complexity of the system. The samples acquired from Bayesian inference can be used in both model selection and in predicting drug response while incorporating the uncertainty of how well the sample represents the population.

Using both PSIS-LOO and WAIC to select the best model is easy to implement and quick to compute if the point-wise log-posterior probability, $\log\left(P\left(x_j|\theta^{(s)}\right)\right)$, for each $x_j \in X^{\text{obs}}$ and $\theta^{(s)} \in S$ were acquired during posterior sampling. Currently, it is not common for Bayesian inference software to record these values during the algorithm, so it does take some time to calculate afterwards.

## 10.1 Selecting the Noise Model

Using the methods outlined in this thesis, constant noise model was found to be the best identifiable model to describe the data. However, this is specific to the data that was analysed in this thesis and this may differ depending on the data acquired in each use case. The noise in the model is simply assumed to be measurement error, however there may be other factors that contributes to this noise term, $\nu\left(t_j\right)$. The inter-individual variability may bias the pooled results, but this was not considered when comparing noise models. Analysing each noise model for each mixed-effects parameter added, however, would have taken a much longer time to analyse. Thus I decided to select the best noise model before adding any mixed-effects parameters, which ensures an identifiable and the simplest sufficient model when doing the mixed-effects analysis. A further study could be done as to whether selecting the best noise model before or after the mixed-effects model significantly affects the results. There may also be environmental factors, such as an unrelated illness, which could affect blood cell levels on a day to day basis. This would require more complexity in the model to incorporate and may be indistinguishable from observation noise when there is a large number of blood cells. This error could also come from imprecision in the sampling or dosing time. It is unlikely that blood samples were acquired at exactly 48 hours after a dose for every rat. Instead, it would be expected that the true sample time for an individual, $t_j^i$ would be randomly distributed around the recorded sample time $t_j$.

The results we have obtained for the experimental data differ to the results acquired by Friberg *et. al.* [17] where they determined that combined noise (where $\eta$

was one) was the most appropriate. However there were several differences between that study and this work. First the paper does not describe in any detail how the choice of noise model was made. If simply the likelihood score was used then even that would not agree with my results which found constant noise had the greatest log-likelihood. Additionally I found that combined noise had identifiability issues, even in the synthetic data, which was not discussed in the previous study. The previous study, however, had much different data to the data I am using. It uses a much larger population of human participants, and observes and models Neutrophil rather than platelet count. A more comparable study would be the paper by Kobuchi *et. al.* [35], which uses a similar model to analyse platelet count in rats. The paper uses Akaike's Information Criteria (AIC) [2] model selection for their assumptions on noise, but does not state which model was selected. It would be expected that the physiological non-drug related PD parameters, $R_0$, $MTT$ and $\gamma$, would be similar with those found in the Kobuchi paper as both our studies model the same system in rats. However there were some discrepancies between these two works. They found $MTT$ as 101.76 hours (CV% 21.4) which is close to my value of 87.65 hours. However my inferred values for $R_0$ and $\gamma$ deviate from theirs ($R_0$: 719 $10^3/\mu L$ with CV% of 4.51. $\gamma$: 0.78 with CV% of 19.3).

## 10.2 Selecting the population model

Selecting the mixed-effects population model is a much more difficult problem due to the greater number of parameters that are added with each decision made. Friberg *et. al.* [17] set $\psi = [R_0, \mathrm{MTT}, s]$ with population PK parameters acquired from previous studies. Whereas we started to see convergence problems with two population parameters. It may be beneficial to again separate PK and PD to simplify the parameter search space. However, in the current software tool for performing Bayesian inference on mixed-effects models, CHI, fixing individual parameter values is not as easy as fixing fixed-effects parameter values. Kobuchi *et. al.* [35] used a fixed-effects population model in their study, as a potential basis for further study into population dynamics within humans.

## 10.3 Future work

This work provides a good basis for future work. The methods outlined above provide a good standardised structure upon which processes can be refined, further analyses

and modelling incorporated, and adaptions to more specific cases be made. I have already mentioned analysis needs to be done on how to chose the order in which to introduce certain features into the model. Also, more development on increasing profile likelihood efficiency can be made. There are other avenues that the above methodology does not cover, which I have outlined below, with suggestions on how they can be incorporated into the analysis.

### 10.3.1 Further Experimentation

In this report I used only one experimental data-set provided by Roche. The clinical trial however recorded results for four different anti-tumour drugs administered and observations on eight types of blood cell. The analysis on multiple datasets like this can be incorporated in one of two ways. The first would be to assume the drugs have no relation to each other and to fit the model to each drug and cell count separately. The second would be to assume that the drug does not affect system-related parameters (such as $(R_0, \mathrm{MTT}, \gamma, \sigma_{\mathrm{PD}})$ for the Friberg model) and these parameters would be the same across all data for a particular blood cell count. Similarly, the blood cell observations would not affect the PK model parameters. This second approach would allow the use of a greater amount of data in determining these system parameters.

One way to improve practical identifiabilty of a model is to reduce the complexity of the model as I have done in this report by using a noise model with a single parameter. An alternative solution is to gather more data until these parameters become identifiable, which is also useful in decreasing the uncertainty of our parameters. As clinical trials and preclinical experiments are expensive especially compared to modelling, it would be beneficial to optimally design experiments such that each new measurement contains the maximum amount of information. Additionally, when analysing the posterior in new trials, previous results can be incorporated into the prior, meaning the data from multiple previous studies can help provide a clearer result. This does require transparency, and open publication of all, even negative, results, to ensure no biases are being introduced using this method.

### 10.3.2 Alternative deterministic models

The form of the noise is not the only assumption made when building the model. For example, the direct drug effect model, was assumed to be linear in this report

represented by Equation 2.7. However, this may not be the case and a non-linear model such as in Equation 2.8 might be better.

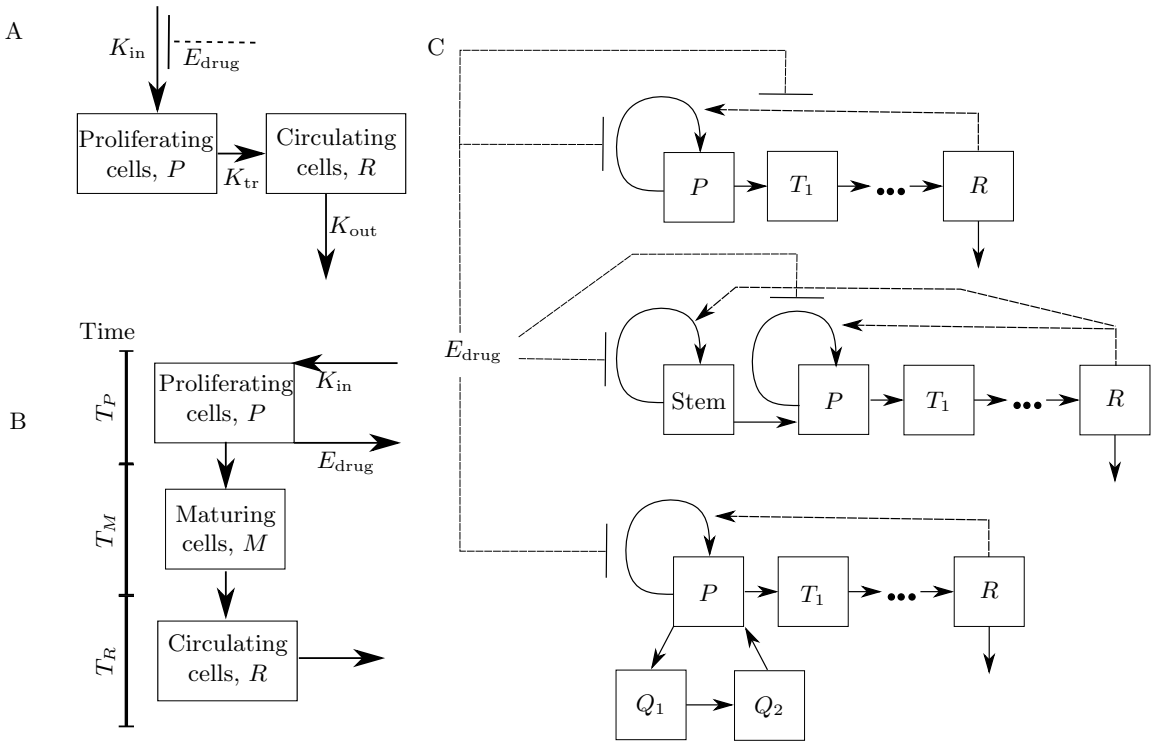



Figure 10.1: A) Minami model of myelotoxicity. A two-compartment model similar to the Dayneka model, however instead of $E_{\text{drug}}$ being defined by the current drug concentration, it is defined by the amount of time the proliferating cells would be exposed to the drug. B) Krzyzanski model of myelotoxicity. A three-compartment model similar to the Zamboni model, however there are no rates of transfer between compartments, instead each cell spends $T_P$, $T_M$, and $T_R$ amount of time in each state. C) Three different models of myelotoxicity proposed by Henrich et al, all extensions of the Friberg Model. The top model introduces different potential drug actions, the middle assumes a pool of slower proliferating stem cells that can balance stabilise the population of faster proliferating cells, and the bottom assumes a population of inactive quiescent cells which can also stabilise the population of faster proliferating cells

Alternatively comparisons with other models of the blood cell life cycle could also be done. In Chapter 2, I described many different models and how they built upon each other to represent the features seen in the data. However, there were other approaches to incorporate these same features such as modelling the drug effect based on the time of exposure to the drug of the proliferating cells. In Minami *et al.*'s paper [40], they take the general PD model proposed by Dayneka *et. al.* [11] and change $E_{\text{drug}}(t)$ to be based on the concentration of drug over the preceding $T_s$ hours, where

$T_s$ is the amount of time that the proliferating cell is sensitive to the drug (see Figure 10.1A). This means

$$E_{\text{drug}}(t) = \frac{AUC(t)}{AUC(t) + EC_{50}},$$

where

$$AUC(t) = \begin{cases} \int_0^t C_0(z)\ \text{dz}, & \text{if } t < T_s \\ \int_{t-T_s}^t C_0(z)\ \text{dz}, & \text{if } t \geq T_s \end{cases}, \tag{10.1}$$

$EC_{50}$ is the value of AUC where there is 50% inhibition, and $C_0(t) = V_c^{-1} A_c(t)$ is the concentration of drug in the central compartment. Minami *et al.* also use a two-compartment model to represent the transfer of cells from the bone marrow into circulation, giving the following system of ODEs,

$$\begin{aligned} \frac{\text{d}P}{\text{d}t} &= K_{\text{in}}(1 - E_{\text{drug}}(t)) - K_{\text{tr}}P, \\ \frac{\text{d}R}{\text{d}t} &= K_{\text{tr}}P - K_{\text{out}}R, \end{aligned}$$

where $K_{\text{in}} > 0$ is the production rate of proliferating cells, $P$, $K_{\text{tr}} > 0$ is the rate of transfer from proliferating cells to circulating cells, $R$, and $K_{\text{out}} > 0$ is the elimination rate of circulating cells.

In Krzyzanski and Jusko's paper, the authors used a cell lifespan model [36]. They separate the life cycle of blood cells into three stages, proliferating, maturing and circulating (see Figure 10.1B) where a cell would spend exactly $T_P$, $T_M$ and $T_R$ hours in each state, respectively. $E_{\text{drug}}$ in this case does not inhibit proliferation but kills proliferating cells. This produces the system of ODEs,

$$\begin{aligned} \frac{\text{d}P}{\text{d}t} &= K_{\text{in}} - K_{\text{in}} \exp\left(-\int_{t-T_P}^t E_{\text{drug}}(z)\ \text{dz}\right) - E_{\text{drug}}(t)\,P, \\ \frac{\text{d}M}{\text{d}t} &= K_{\text{in}} \exp\left(-\int_{t-T_P}^t E_{\text{drug}}(z)\ \text{dz}\right) - K_{\text{in}} \exp\left(-\int_{t-T_P-T_M}^{t-T_P} E_{\text{drug}}(z)\ \text{dz}\right), \\ \frac{\text{d}R}{\text{d}t} &= K_{\text{in}} \exp\left(-\int_{t-T_P-T_M}^{t-T_P} E_{\text{drug}}(z)\ \text{dz}\right) - K_{\text{in}} \exp\left(-\int_{t-T_P-T_M-T_R}^{t-T_P-T_R} E_{\text{drug}}(z)\ \text{dz}\right) \end{aligned}$$

where $E_{\text{drug}}$ can take either the linear or $E_{\max}$ form as shown in Equations 2.7 and 2.8.

These differing model approaches should be compared with each other using the model selection methods described in Chapter 8, to determine whether one particular model approach describes the data better. It may also be beneficial to extend these models to include a negative feedback loop to ensure the rebound seen in the data is sufficiently captured by the model.

The paper by Henrich *et. al.* [23] also extended the Friberg model [17] to include the long term effect of chemotherapy. When a patient receives multiple cycles of chemotherapy the myelotoxic effects become worse and the nadir gets lower at each cycle. The Friberg model or any other model currently explored in this report does not account for this. In these models the circulating blood cells return to base level between cycles with no "memory" of previous cycles. Henrich *et. al.* model several potential systems, as shown in Figure 10.1C that could contribute to the long term "memory" of previous chemotherapy cycles. However, the experimental data that I have access to, and is common in the earlier stages of drug development, does not contain any long-term data. Thus only at a later stage in drug development would it be useful to incorporate and test these additional modelling of long-term effects. However, as it is an extension of the Friberg model, earlier analyses of the Friberg model and previous clinical trials, can be incorporated via the prior into the analysis of the Heinrich models.

# Appendix A

# Software

Most of the above work was developed in jupyter notebooks, these notebooks can be found on GitHub at https://github.com/Rebecca-Rumney/Myleotoxicity-PKPD.git. Both Scipy ODE solvers and Myokit have been used for solving the systems of ODEs. Structural identifiability was determined using Matlab, both using it to help analytically solve equations and using the toolbox STRIKEGOLDD. I have heavily used the PINTS and CHI python packages throughout this thesis for parameter inference, population modelling and optimisation while acquiring the profile likelihood. I have also used optimisers from Scipy and NLOpt for the profile likelihood. For analysing results from Bayesian parameter inference and doing model comparison, I have used the python package Arviz. Graphs have been developed using Plotly or Matplotlib and edited in inkscape.

# Bibliography


[1] L. Aarons, M. O. Karlsson, F. Mentre, F. Rombout, J. L. Steimer, and A. Van Peer. Role of modelling and simulation in Phase I drug development. *European Journal of Pharmaceutical Sciences*, 13(2):115–122, 5 2001.

[2] Hirotugu Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[3] Sara Algeri, Jelle Aalbers, Knut Dundas Morå, and Jan Conrad. Searching for new phenomena with profile likelihood ratio tests. *Nature Reviews Physics 2020 2:5*, 2(5):245–252, 4 2020.

[4] David Augustin. Chi - An open source python package for treatment response modelling, 12 2021.

[5] Nina Baldy, Nicolas Simon, Viktor K. Jirsa, and Meysam Hashemi. Hierarchical Bayesian pharmacometrics analysis of Baclofen for alcohol use disorder. *Machine Learning: Science and Technology*, 4(3):035048, 9 2023.

[6] O. M. Brastein, B. Lie, R. Sharma, and N. O. Skeie. Parameter estimation for externally simulated thermal network models. *Energy and Buildings*, 191:200–210, 5 2019.

[7] René Bruno, Nicole Vivier, Jean Claude Vergniol, Susan L. De Phillips, Guy Montay, and Lewis B. Sheiner. A population pharmacokinetic model for docetaxel (Taxotere): model building and validation. *Journal of pharmacokinetics and biopharmaceutics*, 24(2):153–172, 1996.

[8] Michael Clerx, Martin Robinson, Ben Lambert, Chon Lok Lei, Sanmitra Ghosh, Gary R Mirams, and David J Gavaghan. Probabilistic inference on noisy time series (PINTS). *arXiv preprint arXiv:1812.07388*, 2018.

[9] Richard Creswell, Katherine M. Shepherd, Ben Lambert, Gary R. Mirams, Chon Lok Lei, Simon Tavener, Martin Robinson, and David J. Gavaghan. Understanding the impact of numerical solvers on inference for differential equation models. 7 2023.

[10] Davey Daniel and Jeffrey Crawford. Myelotoxicity From Chemotherapy. *Seminars in Oncology*, 33(1):74–85, 2 2006.

[11] Natalie L Dayneka, Varun Garg, and William J Jusko. Comparison of four basic models of indirect pharmacodynamic responses. *Journal of pharmacokinetics and biopharmaceutics*, 21(4):457–478, 1993.

[12] 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.

[13] Neelima Denduluri, Debra A. Patt, Yunfei Wang, Menaka Bhor, Xiaoyan Li, Anne M. Favret, Phuong Khanh Morrow, Richard L. Barron, Lina Asmar, Shanmugapriya Saravanan, Yanli Li, Jacob Garcia, and Gary H. Lyman. Dose Delays, Dose Reductions, and Relative Dose Intensity in Patients With Cancer Who Received Adjuvant or Neoadjuvant Chemotherapy in Community Oncology Practices. *Journal of the National Comprehensive Cancer Network*, 13(11):1383–1393, 11 2015.

[14] Ronan Duchesne, Anissa Guillemin, Fabien Crauste, and Olivier Gandrillon. Calibration, Selection and Identifiability Analysis of a Mathematical Model of the in vitro Erythropoiesis in Normal and Perturbed Contexts. *In Silico Biology*, 13(1-2):55–69, 1 2019.

[15] Miro J. Eigenmann, Nicolas Frances, Thierry Lavé, and Antje Christine Walz. PKPD modeling of acquired resistance to anti-cancer drug treatment. *Journal of Pharmacokinetics and Pharmacodynamics*, 44(6):617–630, 12 2017.

[16] Ene I. Ette and Paul J. Williams. Population pharmacokinetics II: Estimation methods. *Annals of Pharmacotherapy*, 38(11):1907–1915, 11 2004.

[17] Lena E. Friberg, Anja Henningsson, Hugo Maas, Laurent Nguyen, and Mats O. Karlsson. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *Journal of Clinical Oncology*, 20(24):4713–4721, 12 2002.

[18] Johan Gabrielsson and Daniel Weiner. Non-compartmental Analysis. pages 377–389, 2012.

[19] Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. Inference and assessing convergence. In *Bayesian data analysis*, chapter 11.4. 3rd edition edition, 2014.

[20] Jana L. Gevertz and Irina Kareva. Minimally sufficient experimental design using identifiability analysis. *npj Systems Biology and Applications 2024 10:1*, 10(1):1–14, 1 2024.

[21] Jerome E. Groopman and Loretta M. Itri. Chemotherapy-Induced Anemia in Adults: Incidence and Treatment. *JNCI: Journal of the National Cancer Institute*, 91(19):1616–1634, 10 1999.

[22] Nikolaus Hansen. The CMA evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

[23] Andrea Henrich, Markus Joerger, Stefanie Kraff, Ulrich Jaehde, Wilhelm Huisinga, Charlotte Kloft, and Zinnia Patricia Parra-Guillen. Semimechanistic bone marrow exhaustion pharmacokinetic/pharmacodynamic model for chemotherapy-induced cumulative neutropenia. *Journal of Pharmacology and Experimental Therapeutics*, 362(2):347–358, 2017.

[24] Philip A. Hines, Rosanne Janssens, Rosa Gonzalez-Quevedo, Apolline I.O.M. Lambert, and Anthony J. Humphreys. A future for regulatory science in the European Union: the European Medicines Agency's strategy. *Nature Reviews Drug Discovery 2021 19:5*, 19(5):293–294, 3 2020.

[25] Matthew D Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

[26] Tanin Intragumtornchai, Jutharat Sutheesophon, Pranee Sutcharitchan, and Daratana Swasdikul. A Predictive Model for Life-Threatening Neutropenia and Febrile Neutropenia after the First Course of CHOP Chemotherapy in Patients with Aggressive Non-Hodgkin's Lymphoma. *https://doi.org/10.3109/10428190009089435*, 37(3-4):351–360, 2009.

[27] James R Jacobs. Infusion Rate Control Algorithms for Pharmacokinetic Model-driven Drug Infusion Devices. *International Anesthesiology Clinics*, 33(3), 1995.

[28] Livija Jakaite, Dayou Li, Manuel Alberto, M Ferreira, and Yoon Lee. Bayesian Nonlinear Models for Repeated Measurement Data: An Overview, Implementation, and Applications. *Mathematics 2022, Vol. 10, Page 898*, 10(6):898, 3 2022.

[29] David Janzén, Mats Jirstrand, Neil D. Evans, and Michael Chappell. Structural Identifiability in Mixed-Effects Models: Two different approaches. *IFAC-PapersOnLine*, 48(20):563–568, 1 2015.

[30] David L.I. Janzén, Linnéa Bergenholm, Mats Jirstrand, Joanna Parkinson, James Yates, Neil D. Evans, and Michael J. Chappell. Parameter identifiability of fundamental pharmacodynamic models. *Frontiers in Physiology*, 7(DEC):590, 12 2016.

[31] David L.I. Janzén, Mats Jirstrand, Michael J. Chappell, and Neil D. Evans. Extending existing structural identifiability analysis methods to mixed-effects models. *Mathematical Biosciences*, 295:1–10, 1 2018.

[32] Ross H Johnstone, Eugene T Y Chang, Rémi Bardenet, Teun P De Boer, David J Gavaghan, Pras Pathmanathan, Richard H Clayton, and Gary R Mirams. Uncertainty and variability in models of the cardiac action potential: Can we build trustworthy models? *Journal of Molecular and Cellular Cardiology*, 96:49–62, 2016.

[33] Steven J. Kathman, Daphne H. Williams, Jeffrey P. Hodge, and Mohammed Dar. A Bayesian population PK-PD model for ispinesib/docetaxel combination-induced myelosuppression. *Cancer Chemotherapy and Pharmacology*, 63(3):469–476, 2 2009.

[34] Jaan Kiusalaas. Introduction to Optimization. In *Numerical Methods in Engineering with MATLAB®*, pages 382–410. Cambridge University Press, 2005.

[35] Shinji Kobuchi, Yukako Ito, Taro Hayakawa, Asako Nishimura, Nobuhito Shibata, Kanji Takada, and Toshiyuki Sakaeda. Semi-physiological pharmacokinetic–pharmacodynamic (PK–PD) modeling and simulation of 5-fluorouracil for thrombocytopenia in rats. *http://dx.doi.org/10.3109/00498254.2014.943335*, 45(1):19–28, 1 2014.

[36] Wojciech Krzyzanski and William J Jusko. Multiple-pool cell lifespan model of hematologic effects of anticancer agents. *Journal of pharmacokinetics and pharmacodynamics*, 29(4):311–337, 2002.

[37] Chon Lok Lei, Michael Clerx, David J. Gavaghan, and Gary R. Mirams. Model-driven optimal experimental design for calibrating cardiac electrophysiology models. *Computer Methods and Programs in Biomedicine*, 240:107690, 10 2023.

[38] Tim Litwin, Jens Timmer, and Clemens Kreutz. Optimal Experimental Design Based on Two-Dimensional Likelihood Profiles. *Frontiers in Molecular Biosciences*, 9:37, 2 2022.

[39] Lixoft SAS. Monolix.

[40] Hironobu Minami, Yasutsuna Sasaki, Nagahiro Saijo, Tomoko Ohtsu, Hirofumi Fujii, Tadahiko Igarashi, and Kuniaki Itoh. Indirect-response model for the time course of leukopenia with anticancer drugs. *Clinical Pharmacology & Therapeutics*, 64(5):511–521, 1998.

[41] Diane R Mould and Richard N Upton. Basic concepts in population modeling, simulation, and model-based drug development. *CPT: pharmacometrics & systems pharmacology*, 1(9):1–14, 2012.

[42] Diane R Mould and Richard N Upton. Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development—Part 2: Introduction to Pharmacokinetic Modeling Methods. *CPT: Pharmacometrics & Systems Pharmacology*, 2(4):38, 2013.

[43] Michael C. Neale and Michael B. Miller. The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics*, 27(2):113–120, 1997.

[44] Elisabet I Nielsen and Lena E Friberg. Pharmacokinetic-pharmacodynamic modeling of antibacterial drugs. *Pharmacological Reviews*, 65(3):1053–1090, 2013.

[45] Andreas Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 8 2009.

[46] Lazzaro Repetto. Incidence and clinical impact of chemotherapy induced myelotoxicity in cancer patients: An observational retrospective survey. *Critical Reviews in Oncology/Hematology*, 72(2):170–179, 11 2009.

[47] Edgardo Rivera and Robert E. Smith. Trends in Recommendations of Myelosuppressive Chemotherapy for the Treatment of Breast Cancer: Evolution of the National Comprehensive Cancer Network Guidelines and the Cooperative Group Studies. *Clinical Breast Cancer*, 7(1):33–41, 4 2006.

[48] Christian P Robert and others. *The Bayesian choice: from decision-theoretic foundations to computational implementation*, volume 2. Springer, 2007.

[49] Rodney Rouse, Issam Zineh, and David G. Strauss. Regulatory Science – An Underappreciated Component of Translational Research. *Trends in Pharmacological Sciences*, 39(3):225–229, 3 2018.

[50] Malcolm Rowland, Carl Peck, and Geoffrey Tucker. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annual review of pharmacology and toxicology*, 51:45–73, 2011.

[51] Tarjinder Sahota, Meindert Danhof, and Oscar Della Pasqua. Pharmacology-based toxicity assessment: towards quantitative risk prediction in humans. *Mutagenesis*, 31(3):359–374, 5 2016.

[52] Christoph Scheiermann, Paul S Frenette, and Andres Hidalgo. Regulation of leucocyte homeostasis in the circulation. *Cardiovascular Research*, 107(3):340–351, 2015.

[53] Matthias Schwenkglenks, Ruth Pettengell, Christian Jackisch, Robert Paridaens, Manuel Constenla, André Bosly, Thomas D. Szucs, and Robert Leonard. Risk factors for chemotherapy-induced neutropenia occurrence in breast cancer patients: Data from the INC-EU Prospective Observational European Neutropenia Study. *Supportive Care in Cancer*, 19(4):483–490, 4 2011.

[54] V. Shivva, J. Korell, I. G. Tucker, and S. B. Duffull. An Approach for Identifiability of Population Pharmacokinetic–Pharmacodynamic Models. *CPT: Pharmacometrics & Systems Pharmacology*, 2(6):1–9, 6 2013.

[55] Matthew J. Simpson, Alexander P. Browning, David J. Warne, Oliver J. Maclaren, and Ruth E. Baker. Parameter identifiability and model selection for

sigmoid population growth models. *Journal of Theoretical Biology*, 535:110998, 2 2022.

[56] Vijay K. Siripuram, Daniel F.B. Wright, Murray L. Barclay, and Stephen B. Duffull. Deterministic identifiability of population pharmacokinetic and pharmacokinetic–pharmacodynamic models. *Journal of Pharmacokinetics and Pharmacodynamics*, 44(5):415–423, 10 2017.

[57] Charlotte van Kesteren, Anthe S Zandvliet, Mats O Karlsson, Ron AA Mathôt, Cornelis JA Punt, Jean-Pierre Armand, Eric Raymond, Alwin DR Huitema, Christian Dittrich, Herlinde Dumez, Henri H Roché, Jean-Pierre Droz, Miroslav Ravic, S Murray Yule, Jantien Wanders, Jos H Beijnen, Pierre Fumoleau, and Jan HM Schellens. Semi-physiological model describing the hematological toxicity of the anti-cancer agent indisulam *. *Investigational New Drugs*, 23:225–234, 2005.

[58] Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5):1413–1432, 2017.

[59] Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*, 2015.

[60] D. J. Venzon and S. H. Moolgavkar. A Method for Computing Profile-Likelihood-Based Confidence Intervals. *Applied Statistics*, 37(1):94, 1988.

[61] Linda Wanika, Neil D. Evans, Martin Johnson, Helen Tomkinson, and Michael J. Chappell. In vitro PK/PD modeling of tyrosine kinase inhibitors in non-small cell lung cancer cell lines. *Clinical and Translational Science*, 17(3):e13714, 3 2024.

[62] Sumio Watanabe and Manfred Opper. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010.

[63] Derek Weycker, Mark Hatfield, Aaron Grossman, Ahuva Hanau, Alex Lonshteyn, Anjali Sharma, and David Chandler. Risk and consequences of chemotherapy-induced thrombocytopenia in US clinical practice. *BMC Cancer*, 19(1):1–8, 2 2019.

[64] Franz-Georg Wieland, Adrian L Hauber, Marcus Rosenblatt, Christian Tönsing, and Jens Timmer. On structural and practical identifiability. *Current Opinion in Systems Biology*, 2021.

[65] Ying Wu, Suresh Aravind, Gayatri Ranganathan, Amber Martin, and Luba Nalysnyk. Anemia and thrombocytopenia in patients undergoing chemotherapy for solid tumors: A descriptive study of a large outpatient oncology practice database, 2000–2007. *Clinical Therapeutics*, 31(PART. 2):2416–2432, 1 2009.

[66] William C Zamboni, David Z DArgenio, Clinton F Stewart, Tom MacVittie, Brian J Delauter, Ann M Farese, Douglas M Potter, Nacole M Kubat, David Tubergen, and Merrill J Egorin. Pharmacodynamic Model of Topotecan-induced Time Course of Neutropenia. *Clinical Cancer Research*, 7(8):2301–2308, 2001.

[67] Ci Zhu, Yan Wang, Xicheng Wang, Chunmei Bai, Dan Su, Bing Cao, and Jianming Xu. Profiling chemotherapy-associated myelotoxicity among Chinese gastric cancer population receiving cytotoxic conventional regimens: epidemiological features, timing, predictors and clinical impacts. *Journal of Cancer*, 8(13):2614, 2017.