THE USE OF SIMULATED DATA IN ANALYSING PARAMETER IDENTIFIABILITY FOR IN–HOST COMPARTMENTAL MODELS OF MRNA VACCINE IMMUNE RESPONSES: AN APPLICATION TO IMPROVING SAMPLING OF PHYSICAL PATIENT DATA

AHAD MOOSA

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN APPLIED AND INDUSTRIAL MATHEMATICS

YORK UNIVERSITY
TORONTO, ONTARIO

APRIL 2026

© AHAD MOOSA, 2026

Abstract

In–host mathematical models have become common place in mathematical epidemiology and immunology for the analysis of disease spread, treatment, and prevention. This theory has been extensively applied to modelling the immune responses of patients following vaccine administration by fitting data attained from clinical trials. Often such data is sparse and does not contain observations of all relevant compartments which can impact parameter identifiability. We consider an analysis of parameter identifiability of an in–host vaccine immune response model for LNP mRNA vaccines using simulated data. We employ the use of the software DAISY, for differential algebra analysis, and Monolix, for NMLE modelling that estimates model parameter values through fitting the simulated data. The results indicate that prior knowledge of T–cell and interleukin dynamics allowing for the fixing of parameters on a population level can yield accurate parameter estimates even when a limited number of compartments are observed.

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

Dedicated to my Grandfather

Attai Khan Yousufzai

A doctor who would treat the poor while asking for nothing in return

Acknowledgements

I would like to thank the members and researchers of the Modelling Infection and Immunity Lab (MI2) affiliated with the Centre for Disease Modelling (CDM) at York University for their helpful feedback and guidance throughout the course of this study.

In particular,

My supervisor, Dr. Iain Moyles, for his constant support, encouragement, and advice over the years despite all the delays. His knowledge of the theory and previous research was instrumental in making crucial decisions surrounding the nature of the experiments and the structure of this study. Without him, neither the research project nor this thesis would be possible.

My co–supervisor and head of MI2, Dr. Jane Heffernan, for her helpful suggestions with regards to improving experimental procedures and the classification of results. Her vast knowledge of existing literature in the field was key in establishing a clear picture of the current state of in–host modelling research.

Dr. Chapin Korosec, for his assistance in setting up the various experimental procedures in the Monolix software and his advice on optimising the statistical methods for the particular datasets used in this study. The process for generating the simulated data used in this study stems mainly from his previous work with laboratory data.

The above three are all contributing authors on the published articles that motivated this study.

Thank you as well to all of my graduate student colleagues at MI2 and the Department of Mathematics and Statistics who consulted with me on the results and helped with practicing my presentations.

And finally, thank you to my family, especially my mother, who stood by me all these years and never stopped pushing me to keep going even when I was completely lost on what to do.

Table of Contents

Chapter 1

We begin this thesis with an introduction on the novel research that motivates this study and overview the mathematical theory, methods, and software used in the experiments of the following chapters. First, we provide a summary on the development of mathematical modelling research in epidemiology to study the spread and treatment of diseases within host bodies stemming from compartmental models of disease spread in large populations. From the development of in–host ordinary differential equation (ODE) models of HIV in the 1990s, all the way to the impact of the coronavirus (COVID–19) pandemic in the 2020s bringing about new research into modelling disease spread and vaccine immune responses within–host. We finish with a discussion on the advances in mRNA vaccines during the pandemic and the previous studies which motivate this research (Section 1.1). Within mathematical models in epidemiology, many parameters may exist that have natural biological interpretations which can be linked to physical quantities that are measurable. This leads to a discussion on the theory of parameter identifiability; the ability to accurately estimate parameter values by fitting data points to a given mathematical model. We outline two forms of identifiability. Structural identifiability, stemming from the impact of model structure and observations which is the primary focus of this work, and practical identifiability, stemming from the impact of the particular datasets that are provided. Both of these forms of identifiability can be analysed using computer software which is extensively employed in this study. We overview the software DAISY (Differential Algebra for the Identifiability of SYstems) which uses differential algebra systems to analyse the structural identifiability of models (Section 1.2). We then overview the software Monolix which, given a particular mathematical model and a provided dataset, uses non–linear mixed effects (NLME) modelling to estimate the parameter values of said model by fitting the dataset to the model equations (Section 1.3). Finally, we formally outline the research that forms the remaining chapters of the thesis (Section 1.4).

1.1 Introduction to in–host mathematical modelling

In the 1990s, researchers in mathematical epidemiology began considering adaptations of mathematical models typically used to study disease spread through human communities to the cellular level. A foundational work on the beginnings of the mathematical modelling of disease spread within a particular host body was published by Anderson and May [1]. By translating the widely used S–I–R (Susceptible–Infected–Recovered) compartmental population model to a single body where a population of virions "infects" a population of cells, many of the ecological insights gained from the previous decades of population model research could be applied to disease spread within an individual host where the body could be considered analogous to an ecological environment. Some of the earliest such research using compartmental ordinary differential equation (ODE) models within host involved the spread of HIV. Nowak et al. [2] and Perelson et al. [3] pioneered an S–I–R style structure for HIV infections where uninfected target cells (x) become infected cells (y) through a non–linear interaction with free virus particles (v). Following works such as Ho et al. [4], which demonstrated a rapid decrease in virus concentrations following the administration of newly developed protease inhibitors, the first crude estimates of HIV viral decay half–life caused researchers to realise that HIV infections were more active within the human body than previously anticipated. The uncertainty of the half–life estimates was the result of relatively infrequent sampling and so Perelson et al. [5] performed a similar study with more frequent sampling that allowed for the separation of the infected cell and free virus compartments. This breakthrough confirmed that HIV continuously replicated within the human body rather than being a latent condition and led to the consideration of implementing principles from population epidemiology and ecology such as the basic reproduction number R_0 and the predator–prey mechanism. Nowak and Bangham [6] formalised these concepts for an in–host model of HIV by defining R_0 as the average number of newly infected cells produced by a single infected cell and implementing the predator–prey mechanism through uninfected cells acting as "prey" and the free virus particles acting as "predators". By parameterising the HIV infection with the uninfected cell (x) production rate λ and death rate δ, infected cell (y) production rate β and death rate α, and free virus (v) production rate k and death rate u, the following three compartment in–host model is derived

$$\frac{dx}{dt} = \lambda - \beta xv - \delta x \tag{1.1a}$$

$$\frac{dy}{dt} = \beta xv - \alpha y \tag{1.1b}$$

$$\frac{dv}{dt} = ky - uv \tag{1.1c}$$

Further studies such as Herz et al. [7] implemented the use of delay differential equations to account for the internal delays within the human body that take place with HIV infections. This transition in the 1990s using population modelling insights applied to infectious disease spread within a single body spurred by the advances in HIV treatments laid the foundation of in–host modelling in mathematical epidemiology.

Building on these foundations, many studies emerged throughout the 1990s and 2000s that employed compartmental ODE models to study disease spread and treatment within–host. Neumann et al. [8] adapted the three compartment model to Hepatitis C in order to analyse the effects of interferon–α therapy. Ciupe et al. did the same for Hepatitis B [9] and expanded the number of compartments to consider refractory cells [10]. Heffernan and Keeling [11] combined the use of an SIR style population model and an in–host model for measles vaccinations demonstrating the ability to nest within–host immune response dynamics inside a broader population dynamics framework. Beginning in the 2010s, researchers began to note that, while many standardised in–host models effectively simulated disease spread and treatment at the cellular level through the interaction between cells and virus particles, a gap remained in modelling the internal dynamics of cells which were often crucial to understanding the effectiveness of treatments. To rectify this, Guedj et al. [12] developed a multiscale model for Hepatitis C that simulated both the intracellular and extracellular scales. This led to improvements of Hepatitis C viral half–life estimate and established a structure through which in–host models could simultaneously consider different processes. A landmark review by Ciupe and Heffernan [13] consolidated much of the theoretical research from the previous decades. This included analyses of standardised models for the infection several diseases such as HIV, Hepatitis B and C, tuberculosis, influenza, dengue, and Zika. The researchers also outlined models for various control mechanisms for mitigating and treating diseases through both innate and antibody mediated immune responses as well as drug therapy. Crucially, they highlighted the trade off between model complexity and prediction accuracy of the in–host dynamics. They concluded that, the inclusion of further biological details specific to particular infections or treatments should be justified by real data to reduce mathematical uncertainty. This foundation became increasingly relevant with the emergence of a new global health crisis in the 2020s.

The onset of the coronavirus SARS–CoV–2 (COVID–19) pandemic in late 2019 to early 2020 spawned a new wave of research in mathematical epidemiology. Given the groundwork laid by the previous decades' mathematical and statistical frameworks, researchers worked to identify the virus' behaviour and predict the disease progression within host bodies as well as the

spread throughout a population. At the start of the pandemic, Li et al. [14] and Lauer et al. [15] utilised mathematical and statistical modelling to establish a basic reproduction number estimate of $R_0 = 2.2$ and an incubation period estimate of approximately 5 days for SARS–CoV–2. These parameters became the starting points for many population models early on in the virus' spread worldwide. Around the same time, He et al. [16] used a mathematical framework to analyse samples from infected patients and concluded that between 30%–57% of secondary cases were infected from pre–symptomatic individuals. Such studies were instrumental in informing the initial responses and public heath policies. This research also extended to in–host models with studies such as Néant et al. [17] where the researchers used a target cell–limited compartmental model with distinctions between different types of infectious cells and virus particles. They concluded that reductions in the peak viral load could contribute to reductions in mortality.

Spurred on by the global push towards combating the virus' spread, research was accelerated towards a new form of vaccine as their development, procurement, and distribution became crucial. Messenger ribonucleic acid (mRNA) vaccines have a history also dating back to the 1990s where studies such as Wolff et al. [18] and Martinon et al. [19] demonstrated that mRNA when injected into host bodies could be utilised as a template for antigen production and activate cellular immunity. Prior to the outbreak of the COVID–19 pandemic, Pardi et al. [20] outlined the effectiveness of lipid nano–particles (LNPs) as carriers that can effectively deliver mRNA to target tissues and cell types. The conclusions surrounding the length of the protein production period following mRNA delivery became foundational in establishing parameters for many subsequent in–host mRNA vaccine models during the pandemic. Zhang et al. [21] summarised recent advances in mRNA vaccines and concluded that they were highly flexible and suitable for emerging pandemic infectious diseases. Early on in the pandemic, Krammer [22] reviewed several COVID–19 vaccines in clinical trials, including mRNA vaccines developed by Pfizer–BioNTech and Moderna, and noted that the insights from previously developed vaccines of SARS and MERS allowed for the accelerated development of COVID–19 vaccines within months rather than years. Sahin et al. [23] demonstrated the effectiveness of the Pfizer–BioNTech mRNA vaccine in stimulating strong antibody and T–cell immune responses which became a key point in calibrating later in–host models of mRNA vaccine immune responses.

By nature of the global development of vaccine clinical trials followed by procurement and mass inoculations, a large influx of data from the immune responses of people following vaccine administration was made available for research. From this, studies emerged where researchers fitted such patient data to in–host mathematical models of COVID–19 infection and treatment from which they looked to estimate model parameters for the given datasets. Tan et al. [24] used a Bayesian framework to estimate model parameter values on a population level from fitted patient data of a

large population using Markov chain Monte Carlo (MCMC) sampling. Xu et al. [25] considered a much smaller population that allowed for individual level parameter estimation using a least–squares approach. A natural question that emerges from such studies is, how easy or difficult is it to estimate the model parameters? This leads to the theory of parameter identifiability which motivates the research of this thesis. We look to analyse the difficulty of accurately estimating model parameters from in–host vaccine models while considering factors such as model structure and the sampling frequency of patient data. By doing so, we hope to improve our understanding of the particular kinds of immunological data that are more important to attain in clinical trial sampling and hence paint a better picture of the immune response dynamics following vaccine administration.

The parameter identifiability analysis of this work is motivated by the following studies from which we also derive the mathematical model that is used to fit data. Farhang–Sardroodi et al. [26] established an initial framework by analysing immune responses to the adenovirus based COVID–19 vaccine ChAdOx1–S (AZD1222) developed by the University of Oxford and AstraZeneca. For their analysis, the researchers employed a seven compartment non–linear ODE model on which they fitted real patient data from said vaccine's clinical trials. The study investigated the impact of delaying the time between two vaccine doses and reducing the dosage of the second vaccine motivated by the global supply chain concerns at the time. They concluded that a regimen of a standard dose followed by a lower dose could still provide sufficient protective capacity which would allow for the conservation of vaccine doses. Korosec et al. [27] then adapted this model for the LNP–based mRNA vaccines BNT162b2 manufactured by Pfizer–BioNTech and mRNA–1273 manufactured by Moderna. A key distinction compared to the model of Farhang–Sardroodi was the inclusion of two separate compartments for LNP and vaccinated target cells unlike the single target cell compartment considered in the adenovirus vaccine model. The researchers simultaneously fitted patient datasets from a number of clinical trials and laboratory studies [28–34]. They concluded that individuals in the 18–55 age group maintained a stronger immunity than those aged 70+ and also noted that individual patients lose over 99% of humoral immunity within 8 months of the second dose regardless of which vaccine is administered. This implied an increased viability of booster doses in maintaining an elevated immune response. Moyles et al. [35] then took the LNP–mRNA model of Korosec and, through non–dimensionalisation, performed an asymptotic timescale analysis which decomposed the model solutions into five distinct timescales. By comparing the non–dimensional model to the same laboratory datasets fitted in Korosec, they concluded with a similar age dependence where the vaccine activation takes longer and the antibody peak occurs sooner in patients aged 55+. Acknowledging the timescale analysis of the non–dimensional model, for this study we will consider the full dimensional model outlined at the beginning of Moyles for our analysis of parameter identifiability.

Many such studies that fit vaccinated patient data from clinical trials suffer from the issue of sparse datasets due to the high costs of administering trials and the difficulties associated with frequent physical sampling [36–39]. Such was the case in Korosec et al. [27] where parameter estimates were based on limited sets of vaccinated patient data. To fill this gap, in this study we consider the usage of simulated data in order to represent a hypothetical case where much denser datasets are available. Simulating datasets for mathematical models in epidemiology in order to observe the effects on parameter identifiability has previously been applied in studies concerning both in–host models [40] and population models [41]. We look to apply this idea to modelling the immune response of patients following the administration of mRNA vaccines for COVID–19. Our hope is that through providing simulated datasets that would not be possible in real clinical trials, we can gain a clearer picture of parameter identifiability in such in–host models allowing for inferences to be made on improving the sampling of physical patient data. We now provide an overview of the theory and methodology surrounding parameter identifiability that we will apply in the following chapters.

1.2 Outline of parameter identifiability and DAISY

In the field of mathematical modelling and dynamical systems, parameter identifiability encompasses the theory of analysing and quantifying our ability to estimate the values of model parameters. This concept of identifiability can be further broken down into two main forms; structural identifiability (also known as theoretical identifiability), and practical identifiability. For the purposes of this study, we primarily consider structural identifiability and touch upon practical identifiability in Section 3.6.

We now begin with an outline of structural identifiability. In mathematical biology and epidemiology, the classification of identifiability is typically applied to the parameters of a mathematical model given by a dynamical system of ordinary differential equations (ODEs) [42, 43]. For the purposes of this study, we assume that the equations are autonomous (not dependent on time) which gives

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t),\, \boldsymbol{\Theta}), \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{1.2}$$

Where

- $\mathbf{x}(t) = [x_1(t),\, ...,\, x_n(t)]^T \in \mathbb{R}^n_{\geq 0}$ is the vector of state variables
- $\mathbf{x}_0 = [x_1(0),\, ...,\, x_n(0)]^T \in \mathbb{R}^n_{\geq 0}$ is the vector of initial conditions

- $\boldsymbol{\Theta} = [\theta_1,\ ...,\ \theta_q]^T \in \mathbb{R}^q_{\geq 0}$ is the parameter vector
- $\mathbf{f}(\mathbf{x}(t),\ \boldsymbol{\Theta}) = [f_1(\mathbf{x}(t),\ \boldsymbol{\Theta}),\ ...,\ f_n(\mathbf{x}(t),\ \boldsymbol{\Theta})]^T$ is the parameterised vector field representing the instantaneous rate of change for $\mathbf{x}(t)$ given $\boldsymbol{\Theta}$

It is often the case that the full set of state variables $\{x_1,\ ...,\ x_n\}$ is not observed and instead the measured observations are given by a subset

$$\mathbf{y}(t;\ \boldsymbol{\Theta}) = \mathbf{x}_S(t), \quad S \subseteq \{1,\ ...,\ n\} \tag{1.3}$$

So we have $\mathbf{y}(t;\ \boldsymbol{\Theta}) \in \mathbb{R}^m_{\geq 0},\ m = |S| \leq n$, as the vector of observed state variable outputs given the parameter vector $\boldsymbol{\Theta}$ which may not be fully known. From this, we attain a formal definition of structural identifiability. As a prerequisite, we assume that $\mathbf{y}(t;\ \boldsymbol{\Theta})$ is known exactly for all $t \geq 0$. In other words, the datasets are continuous in time and free of noise [44]. This notion is key to analysing structural identifiability as any resulting lack of identifiability stems from either the model structure or the lack of complete observations. Now given an unknown parameter $\theta_i \in \boldsymbol{\Theta}$, we say that θ_i is *globally structurally identifiable* if

$$\mathbf{y}(t;\ \boldsymbol{\Theta}) = \mathbf{y}(t;\ \widehat{\boldsymbol{\Theta}}), \quad \forall t \geq 0 \quad \Longrightarrow \quad \theta_i = \widehat{\theta}_i \tag{1.4}$$

If this is true for all $\theta_i \in \boldsymbol{\Theta}$, then we say that the mathematical model is globally structurally identifiable; the observation vector $\mathbf{y}(t;\ \boldsymbol{\Theta})$ uniquely determines the parameter vector $\boldsymbol{\Theta}$. For simplicity, we will refer to such parameters and models as just *globally identifiable* moving forward.

If $\mathbf{y}(t;\ \boldsymbol{\Theta})$ does not uniquely determine $\boldsymbol{\Theta}$ and other parameter vectors exist that are determined by it, then we have a slightly looser form of identifiability. Given an unknown parameter $\theta_i \in \boldsymbol{\Theta}$, we say that θ_i is *locally structurally identifiable* if there exists an open neighbourhood $\mathcal{N}(\boldsymbol{\Theta}) \subset \mathbb{R}^q$ within the parameter space such that

$$\mathbf{y}(t;\ \boldsymbol{\Theta}) = \mathbf{y}(t;\ \widehat{\boldsymbol{\Theta}}), \quad \forall t \geq 0, \quad \widehat{\boldsymbol{\Theta}} \in \mathcal{N}(\boldsymbol{\Theta}) \quad \Longrightarrow \quad \theta_i = \widehat{\theta}_i \tag{1.5}$$

That is, within some neighbourhood of the true parameter value, the observation vector $\mathbf{y}(t;\ \boldsymbol{\Theta})$ uniquely identifies θ_i [42, 43]. If this is true for all $\theta_i \in \boldsymbol{\Theta}$, then we say that the mathematical model is locally structurally identifiable. Once again for simplicity, we will refer to such parameters and models as just *locally identifiable* moving forward. Naturally, parameters that do not satisfy either condition are referred to as *non–identifiable*.

Studies within mathematical epidemiology often lack complete observations of the state variables which usually results in a lack of global identifiability for the mathematical models in question. However, while individual parameters may be non–identifiable, functions of parameters may still be uniquely determined from the observed data. Depending on how the individual parameters are coupled within the function terms, prior knowledge of one parameter may lead to another parameter becoming individually identifiable [44]. This leads to the concept of partial identifiability which we explore computationally in 3.6. For an unknown parameter $\theta_i \in \boldsymbol{\Theta}$, we say that θ_i is *partially structurally identifiable* or just *partially identifiable* if for some function of the parameters $r(\boldsymbol{\Theta})$ containing θ_i

$$\mathbf{y}(t;\, \boldsymbol{\Theta}) = \mathbf{y}(t;\, \widehat{\boldsymbol{\Theta}}), \quad \forall t \geq 0 \quad \Longrightarrow \quad r(\boldsymbol{\Theta}) = r(\widehat{\boldsymbol{\Theta}}) \tag{1.6}$$

In other words, the function $r(\boldsymbol{\Theta})$ may be globally (or locally) identifiable while the individual parameter θ_i is identifiable on the condition that other parameters in the function terms are sufficiently known. Examples of this behaviour will be demonstrated in Sections 3.1 and 3.2.

It is often the case that many state variables within a model are not observed which can complicate the algebraic manipulations that yield observed state variable solutions, especially when dealing with a large number of states. This is an issue that pertains to this study and to address it, we introduce the software DAISY (Differential Algebra for the Identifiability of SYstems), developed by researchers at the University of Padova and University of Cagliari in order to perform structural identifiability analyses of ODE models in mathematical biology, epidemiology, ecology, and many other related fields [45]. As the name implies, DAISY employs the theory of differential algebra to express an ODE system with unknown state variables only in terms of the known variables. The methodology behind the application of differential algebra is outlined in Audoly et al. [46] and Saccomani et al. [47]. One feature of DAISY is that it does not require any experimental data to analyse a mathematical model and relies purely on algebraic manipulations through expressing it as a differential algebra system. Hence, it is especially useful for mathematical modelling studies using data from clinical trials and similar types of data collection in the medical field as potential issues with identifiability can be identified prior to any data collection and help inform the sampling such that a clearer picture of the dynamics can be obtained.

A comprehensive overview of the DAISY software including the algorithm through which the differential algebra system is obtained can be found in Bellu et al. [48] while several examples testing the identifiability of models in mathematical biology are outlined in Saccomani et al. [49]. We now provide a brief outline on the process of analysing structural identifiability given the arbitrary mathematical model $d\mathbf{x}/dt$ from equation 1.2 and its observed state vector $\mathbf{y}(t)$ from equation 1.3. First, the equations of $d\mathbf{x}/dt$ are loaded into DAISY as a `.txt` file and each state

variable x_1, ..., x_n is classified as observed or unobserved as specified by $\mathbf{y}(t)$. The parameters θ_1, ..., θ_q from the corresponding parameter vector $\boldsymbol{\Theta}$ are then assigned initial values given by the prime numbers in ascending order, i.e. $\theta_1 = 2$, $\theta_2 = 3$, $\theta_3 = 5$, and so on. DAISY then runs the algorithmic process through which the differential algebra system is attained as a characteristic set of equations expressed only in terms of the observed variables. As we will see in Chapter 3, these equations can be highly complicated and long which necessitates the use of computer software to analyse them. This system is then solved yielding potentially multiple sets of parameter solutions $\{\widehat{\theta}_1, \ ..., \ \widehat{\theta}_q\}$. These solutions can be compared to the initial values allowing us to apply the aforementioned definitions of structural identifiability. Namely

- If $\theta_i = \widehat{\theta}_i$ is a unique differential algebra solution, then θ_i is *globally identifiable*.
- If $\theta_i = \widehat{\theta}_i$ is part of a finite set of non–unique differential algebra solutions, then θ_i is *locally identifiable*.
- If $\widehat{\theta}_i$ is contained within a function of parameters such that $\theta_i = \widehat{\theta}_i$ is a unique differential algebra solution if one or more other parameters are fixed, then θ_i is *partially identifiable*.
- If none of the above are true, i.e. if no solution for $\widehat{\theta}_i$ exists, then θ_i is *non–identifiable*.

These results provide a complete classification of structural identifiability across all parameters in the mathematical model with the only limitation being the hardware used to run the DAISY software. We encounter this limitation in Section 3.2 where the size and complexity of the problem after removing several observed compartments becomes too high for the employed hardware to handle. For completeness, the hardware is listed at the end of Section 1.4.

We briefly discuss the concept of practical identifiability which as mentioned is not covered as extensively in this study. Unlike structural identifiability which considers our ability to estimate parameter values given observed datasets that are continuous in time and noise free, practical identifiability considers what happens when these datasets represent actual experimental data that is finite and more noisy. A parameter can be structurally identifiable but remain practically non–identifiable if the available data is is not enough to distinguish the true parameter value with a high statistical confidence. Various methods exist for quantifying statistical confidence including the Fisher information matrix [42] and the profile likelihood [50–52]. If the available data observations are too sparse, noisy, or insufficiently distributed to allow for a precise estimation of the parameters, it is possible for not just a single parameter vector but a wide range of parameter vectors to describe the observed data equally well. With these classifications of identifiability outlined, we now discuss the statistical methods through which the parameter estimations take place.

1.3 Monolix and non–linear mixed effects modelling

When dealing with research in mathematical biology and epidemiology, it is common for data to be taken from a population of individuals reflecting the same system. For example, in this study we consider the immune system responses of a population of individuals after vaccine administration. However while the system being modelled may be shared across an entire population, the specific responses of individuals may differ greatly. This can complicate the estimation of parameter values when fitting said data to a mathematical model and hence impact identifiability. Non–linear mixed effects (NLME) modelling is a form of statistical modelling that can be used to help describe these intra–individual differences. In an NLME model, parameter values are distinguished between *fixed effects*, which represent the typical parameter value of the population, and *random effects* which represent the difference between an individual's parameter value and the population value. By doing so, the information surrounding parameter values on the population level can be "borrowed" and used to make inferences for individual parameter values especially when data for particular individuals is very limited [53].

Monolix is a specialised software for parameter estimation using NLME modelling originally created by Marc Lavielle at Inria (the French National Institute for Research in Digital Science and Technology) before being further developed for commercial and industrial use by Lixoft [53–55]. The software estimates parameter values through fitting experimental or simulated datasets to mathematical models typically given by ODEs. As such, Monolix is a useful tool for analysing parameter identifiability, both structural and practical, since the provided datasets can be simulated to any desired number of data points. The primary estimation method is the Stochastic Approximation Expectation–Maximisation (SAEM) algorithm, originally outlined by Delyon et al. [56] and then adapted for NLME modelling in Kuhn and Lavielle [57]. We now provide a detailed summary of the entire procedure for estimating model parameters in Monolix.

Consider the mathematical model $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}(t),\, \boldsymbol{\Theta})$ from equation 1.2. Here we consider $\boldsymbol{\Theta} = [\theta_1,\, ...,\, \theta_q]^T$ to be a vector of population parameters for a population of size N with a corresponding population dataset U which we provide to Monolix. This population dataset is composed of individual datasets $U_1,\, ...,\, U_N$ with associated individual parameter vectors $\boldsymbol{\Psi}_i = [\psi_{i1},\, ...,\, \psi_{iq}]^T$ for each U_i. It is assumed that the observed state variables $\mathbf{y}$ are consistent across the entire population. The SAEM algorithm then estimates $\boldsymbol{\Theta}$ and each $\boldsymbol{\Psi}_i$ through an iterative procedure using the conditional distribution. For completeness, we list each dataset and its corresponding parameter estimates under SAEM in Table 1.7.

Dataset	Parameter Estimates
U (Entire Population)	$\boldsymbol{\Theta} = [\theta_1,\ ...,\ \theta_q]^T$
U_1 (Individual 1)	$\boldsymbol{\Psi}_1 = [\psi_{11},\ ...,\ \psi_{1q}]^T$
U_2 (Individual 2)	$\boldsymbol{\Psi}_2 = [\psi_{21},\ ...,\ \psi_{2q}]^T$
.	.
.	.
.	.
U_N (Individual N)	$\boldsymbol{\Psi}_N = [\psi_{N1},\ ...,\ \psi_{Nq}]^T$

Table 1.7: Population and individual datasets with their corresponding parameter estimates under SAEM in Monolix given a population of size N.

First, the population dataset U is loaded as a `.csv` file which specifies the vector of time points $\mathbf{t}$ on which each individual dataset U_i has observed data (technically these time points can differ between individuals however in this study we consider the case where they are all the same). For each time point t, the file specifies the values taken on at that particular time by an individual for the observed state variables $\mathbf{y}$ [58]. Then the mathematical model equations are loaded in a separate `.txt` file which also explicitly defines $\boldsymbol{\Theta}$ and specifies the observed state variables. These variables are then matched with the observed states from the `.csv` file forming what Monolix refers to as the *structural model* [59]. This allows for the initialisation of parameter values for the population. Each population parameter $\theta_i \in \boldsymbol{\Theta}$ is set to a default initial value of 1 which can be manually adjusted to what the user desires [60]. Along with this, the estimation method under SAEM is set for each parameter which by default is the maximum likelihood estimation [61]. The other options are an a posteriori estimation using a Bayesian approach [62] or fixing the parameter outright at the population level.

To complete initialising the population parameter values, Monolix offers its own proprietary auto–initialisation procedure which we refer to as *auto–init*. This is an iteratively run algorithm, which can be initiated as many times as the user desires, that attempts to find good initial population parameter values θ_i that are ideally close to the maximum likelihood estimates to be attained from SAEM. In each iteration, the auto–init algorithm alternates between two different optimisation steps. First, in a global step, the algorithm samples optimal parameter values using a custom least squares method with a cost function minimising the sum of the squared residuals. The resulting set of optimal parameters is then passed as the starting point of a local step, where the algorithm switches to the Nelder–Mead simplex method [61, 63] which we provide further details for later

in this section. The values attained from the simplex algorithm become the new initial population parameter values following which a new iteration of the algorithm can be initialised. The procedure stops when the current iteration does not improve the least squares cost function by more than 1% [64]. Prior to running SAEM, the observed state variables are assigned a posterior probability distribution [65] which by default is a normal distribution and this setting is maintained throughout the study.

Finally, the SAEM algorithm is initialised which begins the iterative procedure with the starting iteration taking the vector of initial population parameters $\boldsymbol{\Theta}^0 = [\theta_1^0,\ ...,\ \theta_q^0]$. Then at each k^{th} iteration, Monolix generates individual parameter vectors $\boldsymbol{\Psi}_i^k$ by sampling from the conditional distribution using a Markov Chain Monte–Carlo (MCMC) method. The distribution considers $\boldsymbol{\Psi}_i$ as a random variable given each individual dataset U_i and the previous population parameters $\boldsymbol{\Theta}^{k-1}$ [61]. The result is a sample of N individual parameter vectors $\boldsymbol{\Psi}_1^k,\ ...,\ \boldsymbol{\Psi}_N^k$ given by

$$\boldsymbol{\Psi}_i^k \sim p(\boldsymbol{\Psi}_i \mid U_i,\ \boldsymbol{\Theta}^{k-1}) \tag{1.8}$$

Under this procedure, the SAEM algorithm is split into the following phases [61]

- Burn–in Phase:

  At the start of the procedure, Monolix samples from the conditional distribution for several iterations (default 5) attaining samples of individual parameter vectors $\boldsymbol{\Psi}_1^k,\ ...,\ \boldsymbol{\Psi}_N^k$ which are then discarded. This is done in order intialise the procedure in a way that prevents bad initial samples from throwing off the final parameter estimates.

- Exploratory Phase:

  Monolix then continues to sample the individual parameter vectors $\boldsymbol{\Psi}_1^k,\ ...,\ \boldsymbol{\Psi}_N^k$ from the conditional distribution for a set number of iterations (default 500). At each iteration, the sample mean becomes the new population parameter vector

$$\boldsymbol{\Theta}^k = \sum_{i=1}^{N} \boldsymbol{\Psi}_i^k \tag{1.9}$$

- Smoothing Phase:

  Monolix then repeats the sampling of the exploratory phase for a set number of iterations (default 200) however this time, the new population parameter vectors are calculated through

averaging all the previous sample means from the smoothing phase

$$\boldsymbol{\Theta}^k = \frac{1}{k}\left[\sum_{i=1}^{N} \boldsymbol{\Psi}_i^1 + \ldots + \sum_{i=1}^{N} \boldsymbol{\Psi}_i^k\right] \tag{1.10}$$

The final iteration gives the output population parameter vector $\boldsymbol{\Theta}$ and its corresponding individual parameter vectors $\boldsymbol{\Psi}_i$ which become the parameter estimates.

Each population parameter $\theta_i \in \boldsymbol{\Theta}$ can optionally be set to have its random effects disabled and Monolix offers multiple settings for dealing with with such non–variable parameters. The default option is no variability at all in which case, instead of SAEM, Monolix uses a variation of the Nelder–Mead simplex algorithm based on MATLAB's `fminsearch` method [61, 66]. This is a direct search method that does not utilise numerical or analytical gradients however it is not guaranteed to converge to a local minimum of the likelihood function [63]. The other settings are adding decreasing variability via random effects and adding variability via random effects only in the exploratory phase which mix the use of the SAEM and simplex algorithms.

Many other settings exist in Monolix for customising the SAEM algorithm, the setup of the structural model, the initialisation of parameter values, and much more. The above setup and methodology for SAEM is consistently applied throughout this study barring Section 2.6 where we consider the application of the Nelder–Mead simplex algorithm. Similar to DAISY, we encounter certain hardware limitations which in this case affect the density of the datasets U_i that can feasibly be used in the SAEM algorithm. As mentioned before, the hardware is listed at the end of Section 1.4 for completeness. For a more comprehensive breakdown of the software with video examples, we direct the reader to the Monolix documentation [55]. Additionally, Traynard et al. [67] functions as a tutorial from researchers at Lixoft outlining the workflow for professional modelling projects in Monolix which showcases the application of some more advanced features. Finally, Heitzman–Breen et al. [68] serves as an excellent example study showcasing the application of simulated datasets from a within–host mathematical model for analysing parameter identifiability in Monolix. The same methodology is applied in this study which we will now summarise.

1.4 Thesis outline

We now provide a formal outline of the research that forms the remaining chapters of this thesis. In Chapter 2, we introduce the in–host mathematical model for mRNA vaccine immune responses on which we perform an analysis of structural identifiability assuming that all compartments are

observed. We demonstrate, using DAISY, that all of the model parameters are identifiable and look to confirm the classification by fitting simulated data to the model in Monolix. We detail the process of simulating virtual patient data from the model in MATLAB allowing us to attain much denser datasets than are typically seen in clinical trials. Using these simulated datasets, we experiment with estimating the model parameter values in Monolix under various different conditions. We do not observe a complete agreement with the results of DAISY suggesting that the density of the datasets is still not enough to capture all of the dynamics. We simplify the model and estimate the parameters again however discrepancies with DAISY still exist. We then experiment with fixing various parameters and modifying the provided datasets leading to improvements in the parameter estimates. We discuss the analytic solutions of the model and highlight why certain parameters may be problematic to accurately estimate. Finally, we make use of the analytic solutions wherever they are available in order to estimate parameters by compartment which yields some of the most accurate estimates.

In Chapter 3, we discuss cases where not all of the model compartments are observed. As a result of this, DAISY indicates that not all of the model parameters are identifiable allowing for us to make inferences on fixing parameters prior to any experimentation in Monolix. The fixing of parameters given these inferences yields highly accurate estimates which serve as an excellent verification of the true power of DAISY and its differential algebra analysis. Notably, unlike Chapter 2, the fixing of certain parameters improving the identifiability of others can be shown more definitively given the classifications of DAISY which we discuss in detail. All of the datasets provided to Monolix in the experiments up to this point represented individual patients. We consider the case where an entire population is provided as a single dataset and compare the effectiveness of the parameter estimates to the individual case. Finally, returning to the individual datasets, we drastically reduce the number of data points in order to represent more realistic samples taken in clinical trials.

Closing out this research in Chapter 4, we discuss conclusions from the experimental results of Chapters 2 and 3 which suggest prioritising sampling certain types of immune response data in clinical vaccine trials. We finish with recommendations of future work to fill in the gaps left by this research that may offer even better insights as to how the sampling of data in trails can be improved.

1.4.1 Hardware

As noted in Sections 1.2 and 1.3, certain procedures in DAISY and Monolix were limited by the available hardware. For completeness, we list the hardware here

- CPU: Intel Core i7–13700H
- GPU: Nvidia RTX 4060 laptop version
- RAM: 32 GB

Chapter 2

In this chapter, we perform an analysis of parameter identifiability on the in–host compartmental lipid–nanoparticle vaccine model which we discussed at the end of Section 1.1 in our outline of in–host vaccine model research following the proliferation of vaccines for COVID–19. Given the highly coupled structure of the model, we assume throughout that all compartments are observed as state variables. This assumption is relaxed in Chapter 3. By doing this, our hope is that we will be able to distinguish between issues relating to structural identifiability and issues relating to practical identifiability. We first formally outline the model, its parameters, and the process stimulating an immune response following vaccine administration. We briefly discuss the analytic solutions for those compartments where they are attainable (Section 2.1). We then perform a differential algebra analysis of the model using the software Differential Algebra for the Identifiability of SYstems (DAISY) outlined in Chapter 1 which identifies the structural identifiability for all parameters (Section 2.2). Following this, we proceed with a parameter estimation analysis. We simulate the model in MATLAB by using randomly generated parameter vectors to produce numerical solution sets called virtual patients (Section 2.3). Then, using the non–linear mixed effects modelling software Monolix also outlined in Chapter 1, we estimate the model parameters for each virtual patient given the simulated datasets and compare them with the known parameter values generated in MATLAB (Section 2.4). These results do not align with the structural identifiability indicated in DAISY and so we estimate each parameter individually to verify which ones match the structural identifiability implied by the differential algebra analysis and which ones are being poorly estimated due to issues with practical identifiability (Section 2.5). We therefore reduce the model by eliminating compartments which are de–coupled from the antibody immune response. We repeat the parameter estimations in Monolix the results of which are improved but still demonstrate discrepancies with those of DAISY and so we experiment with fixing various parameter sets to their known values (Section 2.6). We repeat this process with slightly modified patient datasets that are better suited to the methods in Monolix (Section 2.7). We then outline the analytic solutions of the reduced model in more detail leading to a discussion on why fixing certain parameters improves the estimates of other parameters (Section 2.8). Finally, we take advantage of the analytic solutions by de–coupling the model compartments wherever possible and estimating the parameter values successively by compartment (Section 2.9).

2.1 The lipid nano–particle vaccine model

We begin with a formal definition and overview of the mathematical model used in this identifiability analysis. This in–host compartmental ordinary differential equation (ODE) model was originally published in Farhang–Sardroodi et al. [26] for adenovirus–based COVID–19 vaccines and was later adapted by Korosec et al. [27] for lipid nano–particle (LNP) messenger ribonucleic acid (mRNA) vaccines. Moyles et al. [35] then performed an asymptotic timescale analysis on the adapted model to better understand its parameter dependence which motivates this analysis.

The mathematical model captures the immune response within a given patient following the administration of an LNP mRNA vaccine. The process begins with the injection of a concentration of LNP vaccine (L) which diffuses through a reservoir of inactive target cells. These target cells are then activated becoming vaccinated cells (V) that promote the production of $CD4^+$ T–cells (T) and cytotoxic $CD8^+$ T–cells (C). The $CD4^+$ T–cells subsequently promote the production of plasma B–cells (B), interferon–γ (IFN–γ, F), and interleukin (I). Finally, the plasma B–cells stimulate the production of immunoglobulin G antibody (A). The ODEs for the model, taken directly from Moyles et al. [35], are given by

$$\frac{dL}{dt} = -\mu_{LV}L - \gamma_L L \tag{2.1a}$$

$$\frac{dV}{dt} = \mu_{LV}L - \gamma_V V \tag{2.1b}$$

$$\frac{dT}{dt} = \mu_{VT}V - \gamma_T T \tag{2.1c}$$

$$\frac{dB}{dt} = \mu_{TB}T + \beta_{BI}\left(\frac{I}{I+s_I}\right)B - \gamma_B B \tag{2.1d}$$

$$\frac{dA}{dt} = \mu_{BA}B - \gamma_A A \tag{2.1e}$$

$$\frac{dC}{dt} = \mu_{VC}V + \beta_{CF}\left(\frac{F}{F+s_F}\right)C - \gamma_C C \tag{2.1f}$$

$$\frac{dF}{dt} = \mu_{TF}T - \beta_{FC}FC - \gamma_F F \tag{2.1g}$$

$$\frac{dI}{dt} = \mu_{TI}T - \beta_{IB}IB - \gamma_I I \tag{2.1h}$$

The parameters governing the model dynamics are as follows. Each μ_{ij} is a priming (growth) rate that activates compartment j from compartment i. Conversely, the natural decay rate of compartment i is given by γ_i. The parameters β_{ij} are non–linear autocatalytic activation/inhibition (growth/decay) rates of compartment j on compartment i. Finally, the parameters s_I and s_F are saturation concentrations of I and F respectively. The time is given by t with the vaccine injection occurring at $t = 0$. Moving forward we will refer to equations 2.1a – 2.1h as the *Full LNP Model*.

We can reasonably assume that an isolated individual with no exposure to any diseases will not have an active immune response. Correspondingly, thre steady state of the full LNP model is zero in each compartment. Given this we could assume initial conditions of zero for V, T, B, A, C, F, and I. However more generally, as there are various infections and challenges in our bodies, we allow for the possibility that a patient has mounted an immune response prior to receiving a vaccination and consider non–zero initial conditions for the terminal compartments A, F, and I. Here we look to maintain consistency with Korosec et al. who based their analysis on physical data attained for A and I; hence giving them the non–zero initial conditions $A(0) = A_0$ and $I(0) = I_0$ while maintaining a zero initial condition for F. We assume that the initial concentration of LNP given by $L(0) = L_0$ is scaled such that it is normalised for all vaccine injections resulting in $L(0) = 1$ for our analysis. The compartments L, V, T, B, A, and C are given in arbitrary units (a.u.) while F and I are in picograms per millilitre (pg/ml) [27]. Finally, the time t is given in days. The definitions, units, and initial conditions of the compartments are summarised in Table 2.2.

Variable	Definition	Units	Initial Condition
L	Lipid nano–particles	a.u.	1
V	Vaccinated Cells	a.u.	0
T	$CD4^+$ T–cells	a.u.	0
B	Plasma B–cells	a.u.	0
A	Antibodies	a.u.	A_0
C	$CD8^+$ T–cells	a.u.	0
F	Interferon–γ	pg/ml	0
I	Interleukin	pg/ml	I_0
t	Time	days	0

Table 2.2: Compartmental definitions, units, and initial conditions of the full LNP model in equations 2.1a–2.1h taken directly from Korosec et al. [27]. Here a.u. stands for arbitrary units.

We briefly overview some analytic solutions to the full LNP model to touch upon some suspected issues with identifiablity. A full analysis of the solutions will be shown in Section 2.8. Equations

2.1a (L), 2.1b (V), and 2.1c (T), are a cascade of first order linear ODEs which can be solved successively. Also, A is a first order linear ODE given the unknown solution $B(t)$ and thus is can be written with the general compact integral form. The full solutions are shown in Appendix Awith the final forms as follows

$$L(t) = e^{-(\mu_{LV}+\gamma_L)t} \tag{2.3a}$$

$$V(t) = \frac{\mu_{LV}(e^{-(\mu_{LV}+\gamma_L)t} - e^{-\gamma_V t})}{\gamma_V - \mu_{LV} - \gamma_L} \tag{2.3b}$$

$$T(t) = \frac{\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV} - \gamma_L}\left(\frac{e^{-(\mu_{LV}+\gamma_L)t} - e^{-\gamma_T t}}{\gamma_T - \mu_{LV} - \gamma_L} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V}\right) \tag{2.3c}$$

$$A(t) = e^{-\gamma_A t}\left[A_0 + \mu_{BA}\int_0^t e^{\gamma_A s}B(s)\,ds\right] \tag{2.3d}$$

A quick observation of these solutions already indicates some issues which impact our ability to correctly identify certain parameters. The clearest case of this is in equation 2.3a where the exponential decay rate of LNP is the sum of two parameters $k = \mu_{LV} + \gamma_L$. Therefore, since an arbitrary k can be the sum of an infinite number of two parameter combinations, without prior knowledge of one of μ_{LV} or γ_L, these two parameter values cannot be distinguished when L alone is observed. Of course, this becomes a non–issue when considering observations from both L and V since μ_{LV} is present in both compartments and hence can be identified as the unique growth rate of V. In fact, μ_{LV} is the only parameter that appears in more than one compartment; a peculiarity that is the result of the LNP decaying as a result of priming the target cells via diffusion which is a characteristic not present in downstream compartments. However, in the case where μ_{LV} cannot be correctly identified through observations of V, the identifiability of γ_L becomes less certain given that it is not a unique decay rate.

Due to the cascading effect of the linear ODEs, we note the presence of the following exponential difference terms in equations 2.3b and 2.3c

$$e^{-kt} - e^{-\gamma_V t}, \quad e^{-kt} - e^{-\gamma_T t}, \quad e^{-\gamma_V t} - e^{-\gamma_T t}$$

Here we have once again substituted $k = \mu_{LV} + \gamma_L$. Unlike L, the priming and decay rates for V

and T are unique. Therefore, we do not have the issue of multiple parameters combining into a single observed growth or decay rate which we discussed for equation 2.3a. However, we note that differences in magnitude between the parameters in each exponential term may result in difficulties identifying them if data on particular growth or decay timescales is under–resolved.

2.2 DAISY analysis of the model

We now consider an algebraic analysis of structural identifiability for the full LNP model in DAISY. For this initial testing, we observe all compartments as state variables; considering them as inputs given the full LNP model. In this case, we detail the differential algebra analysis and apply the definition of structural identifiability as outlined in Section 1.2. To begin with, we highlight the fact that given a continuous solution curve, which we interpret as a hypothetical "infinite" dataset, and no prior knowledge of any parameter values, we can identify the scalar coefficients of each model term. Observing the model equations 2.1a – 2.1h, we note that the only parameters not captured by the scalar coefficients are the saturation concentrations s_I and s_F given that they appear in the sums $I+s_I$ and $F+s_F$ in the rational non–linear terms of 2.1d and 2.1f. We therefore scale the model by setting $I = s_I\widetilde{I}$ and $F = s_F\widetilde{F}$ which shifts the saturation concentrations into the scalar coefficients of the non–linear terms in 2.1g and 2.1h. We also combine the coefficients $-\mu_{LV}$ and $-\gamma_L$ in equation 2.1a into $-(\mu_{LV}+\gamma_L)$ as seen in the solution 2.3a. Then by subtracting the derivative terms from both sides of each equation which sets them to zero, we attain the differential algebraic system

$$-(\mu_{LV}+\gamma_L)L - \frac{dL}{dt} = 0 \tag{2.4a}$$

$$\mu_{LV}L - \gamma_V V - \frac{dV}{dt} = 0 \tag{2.4b}$$

$$\mu_{VT}V - \gamma_T T - \frac{dT}{dt} = 0 \tag{2.4c}$$

$$\mu_{TB}T + \beta_{BI}\left(\frac{\widetilde{I}}{\widetilde{I}+1}\right)B - \gamma_B B - \frac{dB}{dt} = 0 \tag{2.4d}$$

$$\mu_{BA}B - \gamma_A A - \frac{dA}{dt} = 0 \tag{2.4e}$$

$$\mu_{VC}V + \beta_{CF}\left(\frac{\widetilde{F}}{\widetilde{F}+1}\right)C - \gamma_C C - \frac{dC}{dt} = 0 \tag{2.4f}$$

$$\mu_{TF}T - s_F\beta_{FC}\widetilde{F}C - s_F\gamma_F\widetilde{F} - s_F\frac{d\widetilde{F}}{dt} = 0 \tag{2.4g}$$

$$\mu_{TI}T - s_I\beta_{IB}\widetilde{I}B - s_I\gamma_I\widetilde{I} - s_I\frac{d\widetilde{I}}{dt} = 0 \tag{2.4h}$$

From equations 2.4a – 2.4h, we can identify four different types of scalar coefficient terms

1. Linear source terms associated with the priming rates μ_{ij} while excluding $-(\mu_{LV}+\gamma_L)$
2. Linear decay terms associated with the decay rates γ_i as well as $-(\mu_{LV}+\gamma_L)$
3. Non–linear source terms with the autocatalytic growth rates given by β_{BI} and β_{CF}
4. Non–linear decay terms with saturation given by $-s_F\beta_{FC}$ and $-s_I\beta_{IB}$

Now given an arbitrary parameter vector $\mathbf{p}$, we can collect the non–unity scalar coefficients and define the coefficient vector $\mathbf{c_p}$. Here we choose to neglect the negative signs on the coefficients since including them does not have any effect. Therefore, the coefficient vector is given by

$$\begin{aligned}\mathbf{c_p}^T = [&\mu_{LV}+\gamma_L,\ \mu_{LV},\ \gamma_V,\ \mu_{VT},\ \gamma_T,\ \mu_{TB},\ \beta_{BI},\ \gamma_B,\ \mu_{BA},\ \gamma_A,\ \mu_{VC},\ \beta_{CF},\ \gamma_C\\ &\mu_{TF},\ s_F\beta_{FC},\ s_F\gamma_F,\ s_F,\ \mu_{TI},\ s_I\beta_{IB},\ s_I\gamma_I,\ s_I]\end{aligned}$$

As mentioned earlier, we assume for the time being in this analysis that all compartments are observed as state variables. Now recalling the definition of structural identifiability outlined in Section 1.2, for a given parameter $p \in \mathbf{p}$ to be structurally identifiable, it must satisfy $\mathbf{c_p} = \mathbf{c_{\widehat{p}}} \Longrightarrow p = \widehat{p}$. Applying this for all the coefficients yields

$$\begin{array}{ccc}
\mu_{LV}+\gamma_L = \widehat{\mu}_{LV}+\widehat{\gamma}_L & \mu_{LV} = \widehat{\mu}_{LV} & \gamma_V = \widehat{\gamma}_V \\
\mu_{VT} = \widehat{\mu}_{VT} & \gamma_T = \widehat{\gamma}_T & \mu_{TB} = \widehat{\mu}_{TB} \\
\beta_{BI} = \widehat{\beta}_{BI} & \gamma_B = \widehat{\gamma}_B & \mu_{BA} = \widehat{\mu}_{BA} \\
\gamma_A = \widehat{\gamma}_A & \mu_{VC} = \widehat{\mu}_{VC} & \beta_{CF} = \widehat{\beta}_{CF} \\
\gamma_C = \widehat{\gamma}_C & \mu_{TF} = \widehat{\mu}_{TF} & s_F\beta_{FC} = \widehat{s}_F\widehat{\beta}_{FC}
\end{array}$$

$$s_F\gamma_F = \widehat{s}_F\widehat{\gamma}_F \qquad s_F = \widehat{s}_F \qquad \mu_{TI} = \widehat{\mu}_{TI}$$
$$s_I\beta_{IB} = \widehat{s}_I\widehat{\beta}_{IB} \qquad s_I\gamma_I = \widehat{s}_I\widehat{\gamma}_I \qquad s_I = \widehat{s}_I$$

We note that since most of the terms in the model are linear with single parameter coefficients, the majority of the parameters satisfy the condition immediately and are structurally identifiable. The parameters where this is not immediately clear are γ_L, γ_F, γ_I, β_{FC}, and β_{IB}. However, we only need to perform some minor algebraic manipulations to their relations given that the rest of the parameters are now known to be structurally identifiable. These manipulations are shown in Appendix A. From this we attain

$$\gamma_L = \widehat{\gamma}_L, \quad \gamma_F = \widehat{\gamma}_F, \quad \gamma_I = \widehat{\gamma}_I, \quad \beta_{FC} = \widehat{\beta}_{FC}, \quad \beta_{IB} = \widehat{\beta}_{IB}$$

Therefore, when providing when all compartments are observed, all parameters are structurally identifiable. Thus we can say that for any arbitrary parameter vector $\mathbf{p}$ of the full LNP model

$$\mathbf{c_p} = \mathbf{c_{\widehat{p}}} \Longrightarrow \mathbf{p} = \widehat{\mathbf{p}}$$

In other words, the full LNP model is globally identifiable when observing all compartments simultaneously. Inputting the full LNP model into DAISY and specifying observations from all compartments yields this exact result which confirms the analysis of this section. The full output files from DAISY can be seen in Appendix B.

2.3 Generation of virtual patients

Having shown that the full LNP model is globally identifiable when all compartments are observed, we are interested in the accuracy and robustness of numerical parameter estimation. For this, we look to generate virtual patients for whom data from the model is simulated and the parameter vector $\mathbf{p}$ is known. We first formally define what we mean by virtual patients and provide a detailed description of the process through which they are generated. We assume that the full LNP model given in equations 2.1a – 2.1h represents the true biological in–host process of the immune response following the administration of a mRNA LNP vaccine. Then using MATLAB, we randomly generate a parameter vector $\mathbf{p}$ and numerically solve the model using said parameters to attain a solution matrix

$$U = \begin{pmatrix} | & | & | & | & | & | & | & | \\ \mathbf{L} & \mathbf{V} & \mathbf{T} & \mathbf{B} & \mathbf{A} & \mathbf{C} & \mathbf{F} & \mathbf{I} \\ | & | & | & | & | & | & | & | \end{pmatrix}$$

The columns of U correspond to the numerical solution vectors for each compartment. More specifically, these are the values that the compartmental solutions take on at different time points which we will elaborate on later in this section. Together, we refer to $P = \{\mathbf{p},\ U\}$ as a *virtual patient* for whom we know the exact model parameters and can attain datasets at any desired resolution. To establish a distribution for generating parameters, we refer back to Korosec et al. [27] where physical patient datasets from a number of different laboratory studies [28–34] were used in Monolix to attain both population and individual parameter estimates under the full LNP model. Moving forward, we will refer to these parameter estimates from Monolix as *fits*. We consider a uniform distribution for each parameter p on the interval $(p_{\min},\ p_{\max}) = (0.9 \times l_{\min},\ 1.1 \times l_{\max})$ where $l_{\min}$ and $l_{\max}$ are respectively the minimum and maximum individual parameter fits across all laboratory datasets. These distributions are summarised in Table 2.5.

We briefly discuss the fact that the parameter fits using laboratory data in Korosec et al. [27] act as the basis for establishing the parameter distributions for the virtual patients used in this study. We recognise that there is no guarantee that the estimated parameter values accurately represent a global distribution that can be applied to any arbitrary patient. This is because the laboratory data only encompassed the antibodies (A) and interleukin (I). Hence, there is less certainty with regards to the parameter distributions in the other compartments as one would need to assume that all of those fits are accurate enough to represent a real population of patients. If this is not the case, then the virtual patients that we generate may not be realistic representations of real patients from whom laboratory data in clinical trials is sampled. For example, the fact that γ_L is distributed around very small values may be due to the parameter not being identifiable given the laboratory data used by Korosec et al. and as a result it was estimated to be near zero. However, we note that our motivation behind this analysis is to analyse the effectiveness of using non–linear mixed effects (NLME) modelling to accurately fit simulated patient data. Therefore, while noting that there are uncertainties in the parameter distributions when it comes to how realistically they represent real patients, we focus more on the idea of optimising the methods and datasets used in the parameter estimation such that arbitrary virtual patients can be accurately fitted.

MATLAB's `rand` function allows for us to generate pseudo–random numbers uniformly on any desired interval using a Mersenne Twister algorithm [69]. Applying this to each interval $(p_{\min},\ p_{\max})$ yields the parameter vector $\mathbf{p}$. An important point to consider is that Monolix outputs

parameter values p rounded to either k decimal places if $|p| \geq 1$ or k significant figures if $|p| < 1$ with the default accuracy being $k = 2$. In order to match this format, we round all generated parameter values to 2 decimal places or 2 significant figures respectively. We then employ MATLAB's numerical ODE solver `ode45` [70] which generates the solution matrix U given $\mathbf{p}$ from $t = 0$ to $t = 365$ days. The solver uses an explicit single step Runge–Kutta method that only requires the previous time step $U(t_{n-1})$ in order to compute $U(t_n)$ [71, 72]. Notably, by only specifying the starting time and ending time, the solver uses an adaptive time–stepping scheme through which the time points that the compartment solutions are evaluated on can differ in order to optimise computation times. In order to apply the same set of time points across all compartments, we use MATLAB's `spline` function [73] which uses cubic spline interpolation to regenerate the solution matrix U on a specified time point vector $\mathbf{t}$. We call these time point vectors *timespans*. Without loss of generality, we will assume that all further mentions of solution matrices U refer to these regenerated solutions with the specified timespans $\mathbf{t}$. One exception to the use of the `ode45` solver is for the IFN–γ concentration (F). Referring to Table 2.5, we see that the linear priming and decay rates $\mu_{TF} \in (167.265,\ 227.887)$ and $\gamma_F \in (112.302,\ 293.667)$ are both very large. In fact they are the largest parameters besides the two saturation concentrations. We hence consider 2.1g as a stiff differential equation for which explicit methods such as the Runge–Kutta method used in `ode45` often result in longer computation times due to a large number of steps and can display artifacts and noise when the solutions are plotted [74, 75]. We confirm that this is in fact the case for plotted solutions of F when using `ode45` where the solution curves exhibit extremely small but regular oscillations which are more clearly visible around the peak. We address this issue of stiffness by instead using MATLAB's `ode23t` solver [76] which uses an implementation of the trapezoidal rule that is better suited for solving moderately stiff ODEs compared to the Runge–Kutta method [77, 78]. Note that this solver also uses an adaptive time–stepping scheme. So we re–solve the model using `ode23t` and extract the solution of F which is then splined using the same timespan $\mathbf{t}$ and inputted as the respective column of the splined solution matrix U. This adjustment yields the desired smooth solution curve for F. We summarise the virtual patient generation process in MATLAB for generating a population of size N

1. Define the uniform distribution for each parameter with the values $(p_{\min},\ p_{\max})$, refer to Table 2.5

2. Generate N parameter vectors $\mathbf{p}_1,\ \ldots,\ \mathbf{p}_N$ using `rand` to draw from each uniform distribution

3. For each $\mathbf{p}_i$, generate the solution matrix U_i' given equations 2.1a – 2.1h using `ode45`

Parameter	Biological Interpretation	Unit	$p_{\min}$	$p_{\max}$	$l_{\min}$	$l_{\max}$
μ_{LV}	Vaccinated cell priming rate	days^{-1}	0.153	4.268	0.17	3.88
μ_{VT}	$CD4^+$ T–cell priming rate	days^{-1}	0.675	29.524	0.75	26.84
μ_{TB}	Plasma B–cell priming rate	days^{-1}	0.063	0.143	0.07	0.13
μ_{BA}	Antibody priming rate	days^{-1}	0.063	1.034	0.07	0.94
μ_{VC}	$CD8^+$ T–cell priming rate	days^{-1}	1.8×10^{-5}	2.86×10^{-5}	2.0×10^{-5}	2.6×10^{-5}
μ_{TF}	IFN–γ priming rate	$(\text{pg/ml})\cdot\text{a.u.}^{-1}\cdot\text{days}^{-1}$	167.265	227.887	185.85	207.17
μ_{TI}	Interleukin priming rate	$(\text{pg/ml})\cdot\text{a.u.}^{-1}\cdot\text{days}^{-1}$	0.0315	28.347	0.035	25.77
γ_L	LNP decay rate	days^{-1}	9.0×10^{-5}	1.43×10^{-4}	1.0×10^{-4}	1.3×10^{-4}
γ_V	Vaccinated cell decay rate	days^{-1}	0.0513	0.143	0.057	0.13
γ_T	$CD4^+$ T–cell decay rate	days^{-1}	0.0396	0.121	0.044	0.11
γ_B	Plasma B–cell decay rate	days^{-1}	0.0234	0.594	0.026	0.54
γ_A	Antibody decay rate	days^{-1}	0.0351	0.0528	0.039	0.048
γ_C	$CD8^+$ T–cell decay rate	days^{-1}	0.00252	0.022	0.0028	0.02
γ_F	IFN–γ decay rate	days^{-1}	112.302	293.667	124.78	266.97
γ_I	Interleukin decay rate	days^{-1}	0.0198	0.055	0.022	0.05
β_{BI}	Interleukin to B–cell inhibition growth rate	days^{-1}	1.917	4.884	2.13	4.44
β_{IB}	B–cell to interleukin inhibition decay rate	$\text{days}^{-1}\cdot\text{a.u.}^{-1}$	0.00117	0.00385	0.0013	0.0035
β_{CF}	IFN–γ to $CD8^+$ T–cell inhibition growth rate	days^{-1}	1.26×10^{-6}	1.54×10^{-6}	1.4×10^{-6}	1.4×10^{-6}
β_{FC}	$CD8^+$ T–cell to IFN–γ inhibition decay rate	$\text{days}^{-1}\cdot\text{a.u.}^{-1}$	1.17×10^{-6}	1.43×10^{-6}	1.3×10^{-6}	1.3×10^{-6}
s_I	Saturation concentration of interleukin	pg/ml	900	1100	1000	1000
s_F	Saturation concentration of IFN–γ	pg/ml	540	660	600	600
A_0	Initial antibody concentration	a.u.	0.468	567.325	0.52	515.75
I_0	Initial interleukin concentration	pg/ml	0.585	259.116	0.65	235.56

Table 2.5: Minimum and maximum parameter values $(p_{\min},\ p_{\max})$ for each parameter uniform distribution with respective laboratory dataset fits $(l_{\min},\ l_{\max})$ from Korosec et al. [27] alongside biological interpretations and units. Note that once again, a.u. refers to arbitrary units.

(a) For 2.1g (F), generate the model solutions using `ode23t` and use the solution for F in the respective column of U_i' in order to handle the stiff equation

4. The U_i''s are generated using the adaptive time–stepping schemes and so given a specified timespan **t**, use `spline` to regenerate each solution matrix with consistent time points giving $U_1,\ ...,\ U_N$

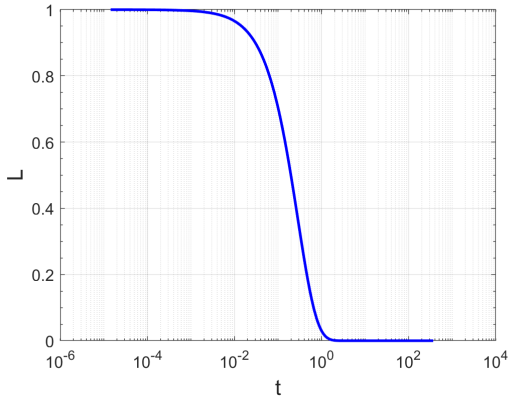
5. Group the parameter vectors and splined solution matrices together as $P_i = \{\mathbf{p}_i,\ U_i\}$ yielding the virtual patients $P_1,\ ...,\ P_N$

For our testing in Monolix, we will occasionally refer to the solution matrices U_i as *datasets* since those are what we consider as patient data which we provide as an input. An example parameter vector is given in Table 2.6 and the corresponding solution plots can be seen in Figure 2.7. With the process of generating simulated patient data outlined, we now consider the experimental setup.

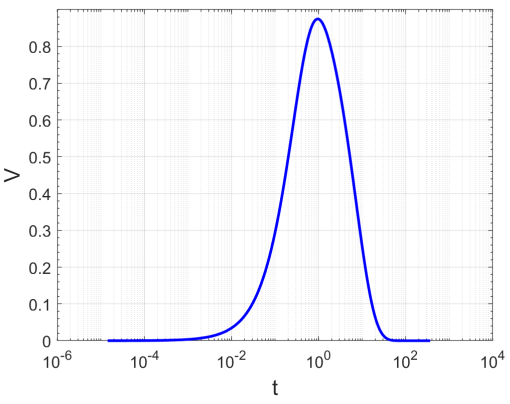
μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{VC}	μ_{TF}	μ_{TI}	γ_L	γ_V	γ_T	γ_B
3.51	26.81	0.073	0.95	2.5×10^{-5}	173.18	7.92	1.2×10^{-4}	0.14	0.12	0.11

γ_A	γ_C	γ_F	γ_I	β_{BI}	β_{IB}	β_{CF}	β_{FC}	s_I	s_F
0.052	0.021	200.33	0.048	2.34	0.0023	1.5×10^{-6}	1.4×10^{-6}	1091.90	618.69

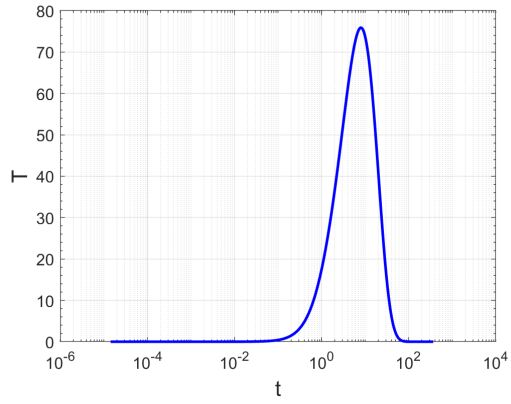
Table 2.6: Parameter vector used for simulating the full LNP model in Figure 2.7. These parameters were generated randomly in MATLAB using the aforementioned procedure.



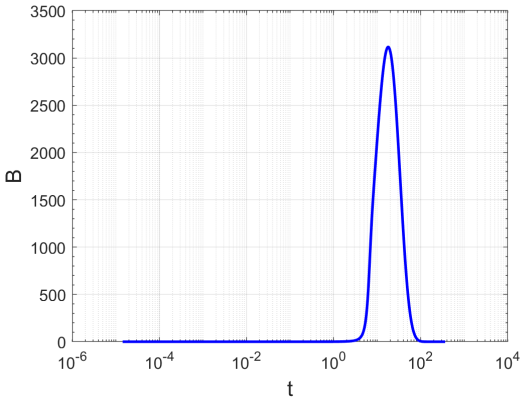
LNP (L)



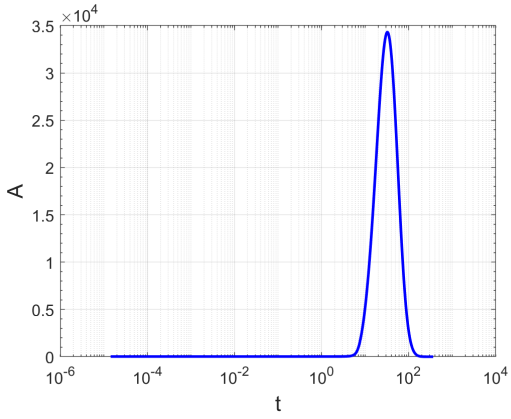
Vaccinated cells (V)



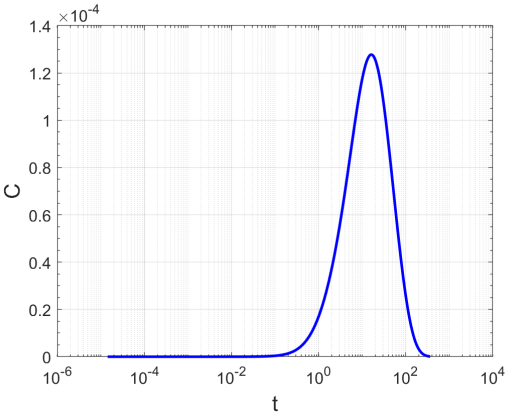
T–cells (T)



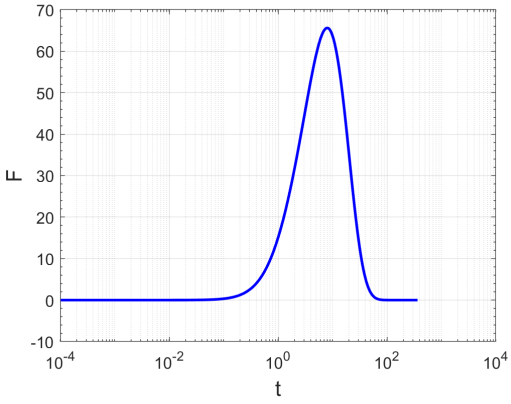
Plasma B–cells (B)



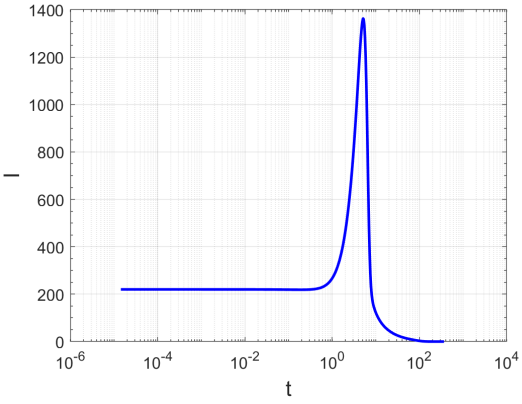
Antibodies (A)



Cytotoxic T–cells (C)



Interferon–γ (F)



Interleukin (I)

Figure 2.7: Example MATLAB plots for each compartment using a set of generated solutions of the full LNP model in equations 2.1a – 2.1h given the parameter vector in Table 2.6. The solutions are generated using `ode45` (except for F which uses `ode23t`) on a set of 10001 linearly spaced time points from the vaccine administration at $t = 0$ to $t = 365$ days. Note that t is plotted logarithmically since several compartments have rapid peak and decay dynamics occurring within the first few days.

2.4 Initial identifiability analysis in Monolix

For an initial experiment in Monolix, we generated five virtual patients in MATLAB from the full LNP model 2.1a – 2.1h given by $P_1^*,\ P_2^*,\ P_3^*,\ P_4^*,\ P_5^*$ where $P_i^* = \{\mathbf{p}_i^*,\ U_i^*\}$. The timespan $\mathbf{t}$ is $10,000$ linearly spaced time points from $t = 0$ to $t = 365$ days. This initial choice on the number of time points was arbitrary and motivated by the need for a large enough dataset that allowed for robustness when fitting parameter values while also minimising concerns regarding practical identifiability. We also considered the impact of dense datasets on computation times which contributed to all such experiments being capped at a maximum of $10,000$ data points. The choice of linear spacing stems from the results of Moyles et al. [35] which indicate varying timescales of growth and decay ranging from on the order of days to on the order of weeks. The full parameter vectors are shown in Table 2.9.

We note that one property of the virtual patient datasets that we provide to Monolix in this experiment is that each individual is fitted separately as its own population of size $N = 1$. We will discuss the case of fiting larger population sizes following the reduction of the model in Section 2.6. Throughout this study, the metric that we use to compare the parameter fits attained in Monolix to the generated parameter values from MATLAB is the percentage error. Formally, given a parameter value p generated in MATLAB and its corresponding fitted value p_f from Monolix, the percentage error is

$$p_\mathrm{e} = \left| \frac{p - p_\mathrm{f}}{p} \right| \times 100\% \tag{2.8}$$

Given a virtual patient P_i, we will refer to the corresponding vector of parameter fits p_f as $\mathbf{f}_i$ and the vector of percentage errors p_e as $\mathbf{e}_i$. For the virtual patients P_i^* in this experiment, this translates to $\mathbf{f}_i^*$ for the fitted parameter vectors and $\mathbf{e}_i^*$ for the error vectors. A formal summary of the experimental procedure in Monolix for each P_i^* is as follows

1. Load the solution dataset U_i^*
2. Set all compartments as observations
3. Load the full LNP model equations
4. Set all parameters as variable and to be fitted under the maximum likelihood estimation with random effects active (Monolix default) [60]
5. Set all initial parameter values to 1 (Monolix default)

6. Set posterior distributions for data on all compartments to normal distributions (Monolix default)

7. Run the SAEM (Stochastic Approximation Expectation–Maximisation) algorithm which outputs the population parameter fits, which are equivalent to the individual patient fits. Recalling the outline of the Monolix parameter fitting procedure in Section 1.3, the setup for SAEM is as follows.

    - 5 iterations in the burn–in phase (dummy iterations that are "thrown–away" in order to get the algorithm properly running)
    - 50–200 iterations in the exploratory phase
    - 25–100 iterations in the smoothing phase

Parameter	$\mathbf{p}_1^*$	$\mathbf{p}_2^*$	$\mathbf{p}_3^*$	$\mathbf{p}_4^*$	$\mathbf{p}_5^*$
μ_{LV}	2.60	3.51	4.00	1.99	1.21
μ_{VT}	8.24	26.81	20.26	19.32	24.93
μ_{TB}	0.11	0.073	0.12	0.12	0.083
μ_{BA}	0.75	0.95	0.78	0.80	0.85
μ_{VC}	2.0×10^{-5}	2.5×10^{-5}	2.2×10^{-5}	2.1×10^{-5}	2.1×10^{-5}
μ_{TF}	174.38	173.18	207.00	208.47	223.60
μ_{TI}	8.43	7.92	4.88	18.59	9.94
γ_L	1.1×10^{-4}	1.2×10^{-4}	1.3×10^{-4}	9.9×10^{-5}	1.0×10^{-4}
γ_V	0.090	0.14	0.054	0.062	0.074
γ_T	0.081	0.12	0.062	0.080	0.090
γ_B	0.072	0.11	0.050	0.57	0.30
γ_A	0.040	0.052	0.037	0.041	0.041
γ_C	0.018	0.021	0.019	0.014	0.019
γ_F	117.60	200.33	238.32	152.90	218.44
γ_I	0.052	0.048	0.031	0.046	0.039
β_{BI}	4.08	2.34	4.74	2.67	4.64
β_{IB}	0.0025	0.0023	0.0013	0.0025	0.0019
β_{CF}	1.4×10^{-6}	1.5×10^{-6}	1.4×10^{-6}	1.5×10^{-6}	1.5×10^{-6}
β_{FC}	1.2×10^{-6}	1.4×10^{-6}	1.3×10^{-6}	1.4×10^{-6}	1.4×10^{-6}
s_I	991.77	1091.90	1053.10	1091.86	976.09
s_F	655.57	618.69	635.42	605.67	608.14
A_0	310.43	20.71	106.40	79.05	43.47
I_0	135.31	220.11	127.20	39.18	14.53

Table 2.9: Parameter vectors $\mathbf{p}_i^*$ of the five virtual patients P_i^* generated from the full LNP model used for the initial experiment in Monolix.

The results of the parameter fits using the SAEM algorithm in Monolix for the initial five virtual patients are shown in are shown in Table 2.10 with the errors shown in Table 2.11. Observing these results, we can assign the model parameters into three categories. We consider a "good"

fit as being one where the error is within 20%, an "average" fit as being one where the error is between 20% and 50%, and a "bad" fit as being one where the error exceeds 50%. The values for the bad fits are in bold. From this, we can label some parameters as consistently having good or average fits such as μ_{LV}, γ_V, and γ_A. We also see that some parameters have good fits with some outliers such as γ_T. Finally, we see that some parameters have consistently bad fits such as β_{CF} and β_{FC}. Overall however, it is difficult to identify a consistent pattern in the errors given the large variation that we see. These results differ greatly from what we observed in the analysis of Section 2.2 which indicates that the model is globally identifiable. This is the first indication of practical identifiability issues with the model. Notably, despite the fact that the datasets used in the experiment are very large and dense given the linear spacing, consistently well fitted parameters are rare. We suspect that this may have something to do with the parameter distributions themselves where the particular values that certain parameters take on make it harder for the algorithm to correctly identify them. We consider this in the next section.

Parameter	$\mathbf{f}_1^*$	$\mathbf{f}_2^*$	$\mathbf{f}_3^*$	$\mathbf{f}_4^*$	$\mathbf{f}_5^*$
μ_{LV}	1.50	3.90	4.79	2.03	1.36
μ_{VT}	24142.68	25.45	26.85	16.05	13.74
μ_{TB}	4.3×10^{-4}	12.61	16.33	3.48	11.40
μ_{BA}	0.45	0.88	0.76	1.60	1.38
μ_{VC}	0.019	1.7×10^{-4}	4.8×10^{-6}	2.8×10^{-4}	1.1×10^{-4}
μ_{TF}	0.27	0.96	4.27	32.60	6.18
μ_{TI}	24477.49	64280.88	1.35	27391.40	42504.28
γ_L	0.59	2.5×10^{-3}	7.9×10^{-5}	3.1×10^{-4}	1.5×10^{-4}
γ_V	0.077	0.15	0.063	0.067	0.080
γ_T	45.41	0.11	0.062	0.081	0.10
γ_B	0.071	0.51	1.14	0.86	0.61
γ_A	0.039	0.052	0.039	0.042	0.041
γ_C	27981.74	0.036	632.18	1.17	0.029
γ_F	0.45	3.72	5.72	21.00	1.51
γ_I	16.35	0.78	4.1×10^{-3}	1.94	1.27
β_{BI}	133.12	0.32	1.07	0.51	0.43
β_{IB}	11.29	23.24	3.3×10^{-4}	6.01	5.40
β_{CF}	15.69	1.12	48.65	0.05	10.22
β_{FC}	8.4×10^{-5}	108.08	0.040	7.35	105.99
s_I	33931.60	0.44	2.2×10^{-3}	9.10	6.34
s_F	82.49	653.72	132590.85	797.23	82063.38
A_0	303.58	25.77	79.25	81.02	49.66
I_0	319.10	200.23	793.22	16.78	4.42

Table 2.10: Fitted parameter vectors $\mathbf{f}_i^*$ of the five initial virtual patients P_i^* from the results of the SAEM algorithm in Monolix.

Parameter	$\mathbf{e}_1^*$	$\mathbf{e}_2^*$	$\mathbf{e}_3^*$	$\mathbf{e}_4^*$	$\mathbf{e}_5^*$
μ_{LV}	42.33	11.25	19.86	2.19	12.15
μ_{VT}	**292909.81**	5.06	32.55	16.93	44.88
μ_{TB}	**99.61**	**17136.44**	**13109.92**	**2806.07**	**13578.48**
μ_{BA}	40.29	7.36	3.13	**101.06**	**61.65**
μ_{VC}	**93263.73**	**588.18**	**78.34**	**1238.06**	**434.46**
μ_{TF}	**99.85**	**99.45**	**97.94**	**84.36**	**97.24**
μ_{TI}	**290191.93**	**811802.39**	**72.33**	**147316.75**	**427445.42**
γ_L	**551842.21**	**2001.11**	38.00	**214.34**	49.37
γ_V	14.63	7.83	16.20	7.70	7.64
γ_T	**56003.52**	6.89	0.23	1.04	11.43
γ_B	1.66	**350.00**	**2191.67**	**50.61**	**107.87**
γ_A	1.88	0.53	5.92	2.13	0.78
γ_C	**154392435.00**	**70.09**	**3405866.35**	**8304.56**	**55.04**
γ_F	**99.62**	**98.14**	**97.60**	**86.26**	**99.31**
γ_I	**31045.43**	**1526.02**	**86.76**	**4095.08**	**3143.91**
β_{BI}	**3159.64**	**86.31**	**77.41**	**80.93**	**90.73**
β_{IB}	**455238.87**	**1010194.12**	**73.86**	**237828.88**	**278818.62**
β_{CF}	**1103385501.00**	**73858750.41**	**3518100602.00**	**3434575.77**	**694285801.40**
β_{FC}	**6719.88**	**7854800020.00**	**3151478.64**	**524387547.90**	**7759323589.00**
s_I	**3321.32**	**99.96**	**100.00**	**99.17**	**99.35**
s_F	**87.42**	5.66	**20766.52**	31.63	**13394.19**
A_0	2.21	24.42	25.52	2.49	14.25
I_0	**135.82**	9.03	**523.58**	0.57	**69.59**

Table 2.11: Percentage error vectors $\mathbf{e}_i^*$ of the five initial virtual patients P_i^* from initial experiment in Monolix. Errors exceeding 50% are in bold

2.5 Identifiability analysis of individual parameters

Given the inconclusive results of Section 2.4, in order to better analyse the effects of the individual parameter distributions on the fits, we consider fitting each parameter individually with all others fixed to their exact values. Our motivation stems from the differential algebra analysis in Section 2.2. We know that the model is globally identifiable and that the large datasets should ideally limit issues with practical identifiability. We wish to confirm whether or not each individual parameter is identifiable on its own using SAEM. Therefore, we study them one–by–one with all other parameters fixed. If a given parameter $p \in \mathbf{p}$ is fitted accurately in the absence of any other variable parameters, then the results serve as a good verification of structural identifiability. We generate ten new virtual patients given by $P_1^\dagger$, ..., $P_{10}^\dagger$ with parameter vectors $\mathbf{p}_i^\dagger$, fitted parameter vectors $\mathbf{f}_i^\dagger$, and percentage error vectors $\mathbf{e}_i^\dagger$. The experimental procedure in Monolix is mostly the same as Section 2.4 but this time with all parameters except one fixed and some minor adjustments to the algorithms which are documented below. For each individual patient

1. Load the solution dataset $U_i^\dagger$

2. Set all compartments as observations

3. Load the full LNP model equations

4. Fix all parameter values to their exact values from MATLAB and disable random effects

5. Set posterior distributions for data on all compartments to normal distributions (Monolix default)

6. For each parameter one at a time

   (a) Unfix the current parameter and select the maximum a posteriori estimation method which is a modification of the maximum likelihood estimation using a Bayesian approach given existing parameter knowledge [62]

   (b) Keep the initial parameter value of the current parameter at the exact MATLAB value

   (c) Re–enable random effects on the current parameter

   (d) Run the SAEM algorithm which outputs the population parameter fits, which are equivalent to the individual patient fits

   (e) Re–fix the current parameter and re–disable its random effects

The percentage errors of the individual parameter fits are shown in Table 2.12.

Parameter	$\mathbf{e}_1^\dagger$	$\mathbf{e}_2^\dagger$	$\mathbf{e}_3^\dagger$	$\mathbf{e}_4^\dagger$	$\mathbf{e}_5^\dagger$	$\mathbf{e}_6^\dagger$	$\mathbf{e}_7^\dagger$	$\mathbf{e}_8^\dagger$	$\mathbf{e}_9^\dagger$	$\mathbf{e}_{10}^\dagger$
μ_{LV}	0.25%	0.00%	0.98%	0.00%	0.00%	0.75%	0.50%	0.98%	1.04%	0.60%
μ_{VT}	0.46%	0.20%	0.39%	0.44%	0.00%	0.25%	0.22%	0.23%	0.31%	0.23%
μ_{TB}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
μ_{BA}	0.00%	0.00%	0.97%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
μ_{VC}	26.92%	85.60%	79.23%	10.71%	3.57%	0.00%	82.63%	48.00%	68.42%	4.17%
μ_{TF}	4.53%	19.78%	20.67%	10.81%	1.08%	34.04%	11.27%	2.65%	32.06%	15.63%
μ_{TI}	0.27%	0.00%	0.42%	0.00%	0.08%	0.15%	0.00%	0.14%	0.27%	0.21%
γ_L	4361.54%	336.36%	4735.16%	1445.45%	218.18%	1200.00%	3400.00%	2130.77%	5263.64%	4614.29%
γ_V	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_T	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_B	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_C	6.25%	45614.29%	11.11%	120.00%	32218.18%	14300.00%	15.79%	464.29%	4.55%	7.69%
γ_F	53.72%	51.04%	136.21%	114.08%	1.05%	164.64%	65.65%	70.98%	399.51%	47.16%
γ_I	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
β_{BI}	0.00%	0.00%	0.44%	0.22%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
β_{IB}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
β_{CF}	3471.43%	976.92%	5257.14%	1033.33%	1900.00%	99.51%	4542.86%	292.31%	42.86%	26.67%
β_{FC}	592.86%	3483.33%	733.33%	138.46%	29130.77%	23.08%	25.00%	21.43%	14066.67%	816.67%
s_I	0.05%	0.05%	0.39%	0.26%	0.02%	0.09%	0.04%	0.00%	0.03%	0.06%
s_F	99.91%	96.49%	94.83%	80.02%	99.80%	31063.36%	100.00%	18.40%	10.54%	20.59%
A_0	0.02%	0.12%	1.07%	0.02%	0.08%	0.14%	0.03%	0.25%	0.26%	0.04%
I_0	0.04%	0.07%	6.51%	0.12%	0.16%	0.11%	1.41%	0.11%	0.23%	0.46%

Table 2.12: Percentage error vectors $\mathbf{e}_i^\dagger$ of virtual patients $P_i^\dagger$ generated from the full LNP model. Each parameter is fitted individually with all other parameters fixed. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Errors exceeding 100% all have the same shade with a new shade at 1000% to indicate extremely divergent fits.

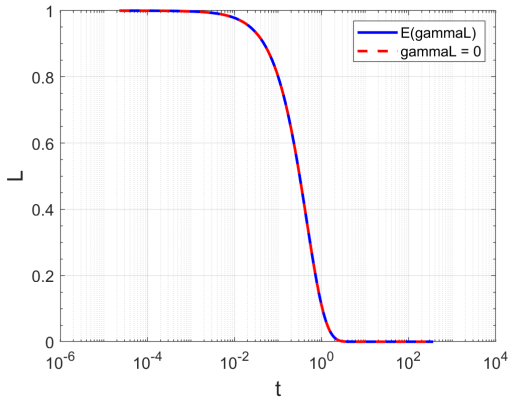
This time the accuracy of the fits is much more promising. Most parameters are consistently fitted exactly or with very small errors that do not exceed 10%. The parameters still displaying large errors are γ_L and all parameters within the C and F compartments which we can now analyse. For γ_L, we can go back to the solution in equation 2.3a in Section 2.1 where the exponential decay stems from the sum $\mu_{LV} + \gamma_L$. We recall that when data from both L and V is observed, both of these parameters can be identified due to the presence of μ_{LV} in equation 2.3b. Yet despite both compartments being observed as well as μ_{LV} being fixed along with all other parameters, γ_L is still not correctly identified. We suspect then that the errors observed with γ_L relate to the difference in magnitudes between the two parameters. Referring to Table 2.5, $\min(\mu_{LV}) \sim \mathcal{O}(10^{-1})$ while $\max(\gamma_L) \sim \mathcal{O}(10^{-4})$. Hence even with the exact value of μ_{LV} known and a very dense dataset, we suspect that the SAEM algorithm in Monolix cannot isolate the minuscule effect on LNP decay from natural processes compared to the decay from priming target cells.

This behaviour with the fits of γ_L of course reflects back to the discussion in Section 2.3. We noted that the choice of basing the parameter distributions on the results of Korosec et al. [27] may not have accounted for certain parameters being non–identifiable from the laboratory data that was used. The fitted values could be pushed towards zero resulting in the distribution sampling around very small numbers which are much harder to identify. We can get an idea of the impact, or lack thereof, of a small γ_L on the immune response by comparing the plots of L and V using the expected values (midpoints) of the relevant parameter distributions versus the same plots but with $\gamma_L = 0$. The parameters are μ_{LV}, γ_L, and γ_V with the following expected values

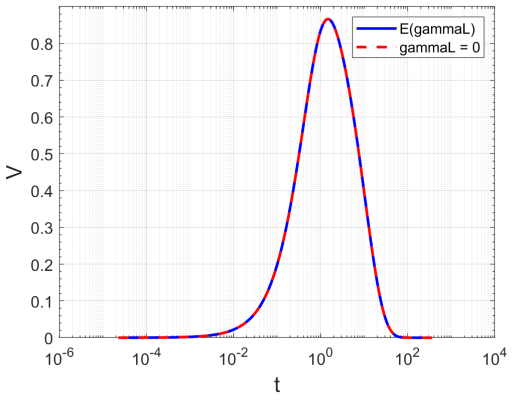
$$E(\mu_{LV}) = 2.2105, \quad E(\gamma_L) = 1.165 \times 10^{-4}, \quad E(\gamma_V) = 0.09715 \tag{2.13}$$

We plot the solutions of L and V given these parameter values compared to the case where $\gamma_L = 0$ in Figure 2.14. It is clear from the plots that the two solutions are indistinguishable due to the fact that γ_L is so small. This unsurprisingly is reflected by the large errors of γ_L in Table 2.12 as the particular values that γ_L takes cannot be distinguished. Since we are more interested in the immune response stimulated by the vaccine rather than the natural decay, we set $\gamma_L = 0$ under the assumption that continued advances in the vaccine engineering process effectively reduce the natural decay rate of LNP to be negligible allowing us to absorb the parameter into the priming rate.

We suspect that there is a similar issue with magnitude for parameters in the C and F compartments. Once again we have the presence of extremely small parameters with $\mu_{VC} \sim \mathcal{O}(10^{-5})$ and $\beta_{CF},\ \beta_{FC} \sim \mathcal{O}(10^{-6})$. However since equations 2.1f and 2.1g are non–linear, we are unable to attain analytic solutions. For this we can refer back to the discussion in Section 2.3 regarding the



LNP (L)



Vaccinated cells (V)

Figure 2.14: MATLAB plots for L and V comparing the solutions with the expected parameter values from equation 2.13 in blue to the solutions where $\gamma_L = 0$ in red. For both compartments, the two curves are indistinguishable due to the fact that $E(\gamma_L)$ is very small.

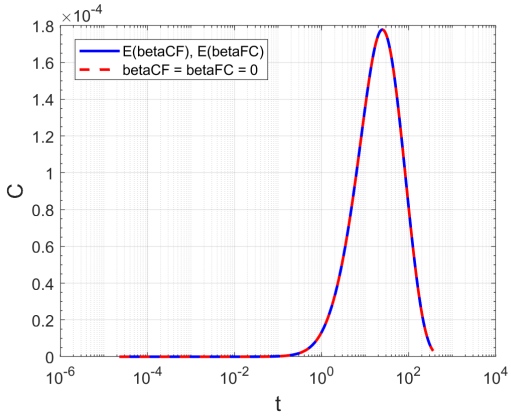
use of fitted parameter data form Korosec et al. [27] as the basis for the parameter distributions. Since the study only observed laboratory data for A and I, we recognise the possibility that these parameters were in fact not identifiable with such a limited dataset and happened to be fitted close to zero. However as these fitted values define the distributions, we cannot distinguish between a true zero and a non–identifiable parameter. Similar to Figure 2.14, we can compare the plots of C and F where the parameters take on their expected values versus where $\beta_{CF} = \beta_{FC} = 0$. We could also set $\mu_{VC} = 0$ however then there would be no priming of C and so the curve would be flat. For this we must also generate solutions to the upstream compartments L, V, and T which means that the parameters present within them must be defined as well. Therefore, the expected values that we consider are

$$E(\mu_{LV}) = 2.2105, \quad E(\mu_{VT}) = 15.0995, \quad E(\mu_{VC}) = 2.33 \times 10^{-5}, \quad E(\mu_{TF}) = 197.576 \quad (2.15)$$

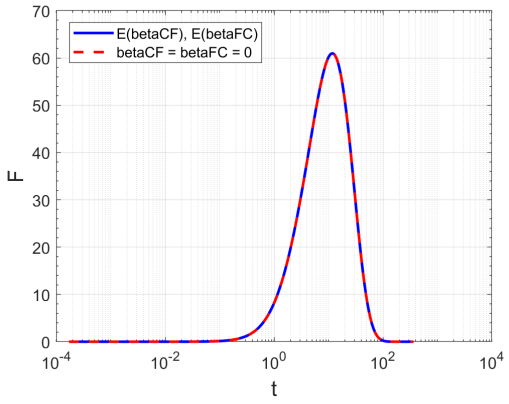
$$E(\gamma_L) = 1.165 \times 10^{-4}, \quad E(\gamma_V) = 0.09715, \quad E(\gamma_T) = 0.0803, \quad E(\gamma_C) = 0.01226$$

$$E(\gamma_F) = 202.9845, \quad E(\beta_{CF}) = 1.4 \times 10^{-6}, \quad E(\beta_{FC}) = 1.3 \times 10^{-6}, \quad E(s_F) = 600$$

The plots are shown in Figure 2.16. Just like before, we see that the solution curves with the expected values are indistinguishable from the solution curves with $\beta_{CF} = \beta_{FC} = 0$ and so similar to the case of γ_L, the large errors in Table 2.12 are not surprising. Luckily, the full LNP model is structured so that the antibody pathway is completely decoupled from C and F. As the antibody dynamics are what we are most interested in, we can choose to neglect the C and F compartments.



Cytotoxic T–cells (C)



Interferon–γ (F)

Figure 2.16: MATLAB plots for C and F comparing the solutions with the expected parameter values from equation 2.15 in blue to the solutions where $\beta_{CF} = \beta_{FC} = 0$ in red. For both compartments, the two curves are indistinguishable due to the fact that $E(\beta_{CF})$ and $E(\beta_{FC})$ are very small.

For completeness before eliminating the compartments, we repeat the experiment with linearised equations by setting $\beta_{CF} = \beta_{FC} = 0$ (note that this also eliminates s_F). This allows us to at the very least attain analytic solutions for the linear dynamics and attempt to infer potential issues with identifiability based off them in the absence of non–linear dynamics. For clarity, we re–write the full LNP model with $\gamma_L = \beta_{CF} = \beta_{FC} = 0$ and the new variables $\widetilde{L}$, $\widetilde{C}$, and $\widetilde{F}$

$$\frac{d\widetilde{L}}{dt} = -\mu_{LV}\widetilde{L} \tag{2.17a}$$

$$\frac{dV}{dt} = \mu_{LV}\widetilde{L} - \gamma_V V \tag{2.17b}$$

$$\frac{dT}{dt} = \mu_{VT}V - \gamma_T T \tag{2.17c}$$

$$\frac{dB}{dt} = \mu_{TB}T + \beta_{BI}\left(\frac{I}{I+s_I}\right)B - \gamma_B B \tag{2.17d}$$

$$\frac{dA}{dt} = \mu_{BA}B - \gamma_A A \tag{2.17e}$$

$$\frac{d\widetilde{C}}{dt} = \mu_{VC}V - \gamma_C\widetilde{C} \tag{2.17f}$$

$$\frac{d\widetilde{F}}{dt} = \mu_{TF}T - \gamma_F\widetilde{F} \tag{2.17g}$$

$$\frac{dI}{dt} = \mu_{TI}T - \beta_{IB}IB - \gamma_I I \tag{2.17h}$$

The initial conditions are the same with $\widetilde{L}(0) = 1$, $\widetilde{C}(0) = 0$, and $\widetilde{F}(0)$. We now regenerate the solution matrices $U_i^\dagger$ with $\gamma_L = \beta_{CF} = \beta_{FC} = 0$ while all other parameter values are kept the same. We then repeat the process in Monolix to fit each parameter individually. The results of these parameter fits are shown in Table 2.18. We still observe large errors in parameters from the $\widetilde{C}$ and $\widetilde{F}$ compartments in the same way as before. We therefore look to the analytic solutions of the linearised equations. Equation 2.17f is of the same form as equation 2.1c and has the same initial condition $\widetilde{C}(0) = 0$. Therefore, without loss of generality, we can write the solution $\widetilde{C}(t)$ in the same way as equation 2.3c while replacing μ_{VT} with μ_{VC} and γ_T with γ_C as well as setting $\gamma_L = 0$.

Parameter	$\mathbf{e}_1^\dagger$	$\mathbf{e}_2^\dagger$	$\mathbf{e}_3^\dagger$	$\mathbf{e}_4^\dagger$	$\mathbf{e}_5^\dagger$	$\mathbf{e}_6^\dagger$	$\mathbf{e}_7^\dagger$	$\mathbf{e}_8^\dagger$	$\mathbf{e}_9^\dagger$	$\mathbf{e}_{10}^\dagger$
μ_{LV}	0.50%	0.00%	0.65%	0.88%	0.00%	1.13%	1.00%	1.47%	1.04%	0.60%
μ_{VT}	0.28%	0.20%	0.39%	0.28%	0.08%	0.25%	0.16%	0.34%	0.37%	0.19%
μ_{TB}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
μ_{BA}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
μ_{VC}	93.46%	4.00%	26.92%	10.71%	17.86%	0.00%	66.84%	12.00%	71.58%	4.17%
μ_{TF}	8.92%	13.77%	4.48%	17.19%	1.09%	34.85%	29.61%	2.93%	22.16%	21.35%
μ_{TI}	0.67%	0.00%	0.09%	0.14%	0.00%	0.15%	0.00%	0.19%	0.27%	0.15%
γ_V	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_T	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_B	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
γ_C	12.50%	23257.14%	511.11%	5.00%	41900.00%	1020.00%	5.26%	7.14%	4.55%	23.08%
γ_F	50.57%	62.23%	152.46%	78.92%	1.08%	72.75%	44.52%	69.94%	399.88%	69.01%
γ_I	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.50%	0.00%
β_{BI}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
β_{IB}	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
s_I	0.02%	0.15%	0.17%	0.11%	0.03%	0.13%	0.08%	0.08%	0.07%	0.03%
A_0	0.83%	0.05%	0.12%	0.06%	0.07%	0.81%	0.46%	0.03%	0.19%	0.10%
I_0	0.09%	0.14%	0.26%	0.26%	0.14%	0.04%	0.77%	0.00%	0.41%	0.58%

Table 2.18: Percentage error vectors $\mathbf{e}_i^\dagger$ of virtual patients $P_i^\dagger$ generated from the modified full LNP model with $\gamma_L = \beta_{CF} = \beta_{FC} = 0$. Each parameter is fitted individually with all other parameters fixed. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Errors exceeding 100% all have the same shade with a new shade at 1000% to indicate extremely divergent fits.

$$\widetilde{C}(t) = \frac{\mu_{VC}\mu_{LV}}{\gamma_V - \mu_{LV}}\left(\frac{e^{-\mu_{LV}t} - e^{-\gamma_C t}}{\gamma_C - \mu_{LV}} - \frac{e^{-\gamma_V t} - e^{-\gamma_C t}}{\gamma_C - \gamma_V}\right) \tag{2.19a}$$

We can also now solve equation 2.17g by continuing the linear cascade of solutions from equation 2.3c and once again with $\gamma_L = 0$. Since this is the fourth solution in the cascade, the terms are more complex and so we present only the final solution in its closed form. The full solution is shown in Appendix A.

$$\widetilde{F}(t) = \frac{\mu_{TF}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left[\frac{e^{-\mu_{LV}t} - e^{-\gamma_F t}}{(\gamma_T - \mu_{LV})(\gamma_F - \mu_{LV})} - \frac{e^{-\gamma_T t} - e^{-\gamma_F t}}{(\gamma_T - \mu_{LV})(\gamma_F - \gamma_T)} \right. \\ \left. - \frac{e^{-\gamma_V t} - e^{-\gamma_F t}}{(\gamma_T - \gamma_V)(\gamma_F - \gamma_V)} + \frac{e^{-\gamma_T t} - e^{-\gamma_F t}}{(\gamma_T - \gamma_V)(\gamma_F - \gamma_T)} \right] \tag{2.19b}$$

Referring to equations 2.19a and 2.19b, it is difficult to conclusively say where the problem lies. The small sizes of μ_{VC}, and γ_C in Table 2.5, being $\mathcal{O}(10^{-5})$ and $\mathcal{O}(10^{-2})$ respectively, may contribute towards non–identifiability in a similar way to the parameters tested in Figures 2.13 and 2.15. Certainly we can imagine that μ_{VC} would be difficult to isolate from the product term $\mu_{VC}\mu_{LV}$ even given a highly dense data set. Similarly, since γ_C is strictly less than μ_{LV} and γ_V, its decay term may be subdominant in the exponential differences $e^{-\mu_{LV}t} - e^{-\gamma_C t}$ and $e^{-\gamma_V t} - e^{-\gamma_C t}$. For the F compartment the inconsistencies are unclear. Especially given that the linearised solutions of both B and I, attained from setting $\beta_{BI} = \beta_{IB} = 0$, have the same form as 2.19b yet the parameters from those compartments were fitted perfectly. The priming and decay rates μ_{TF} and γ_F are both the largest parameters within their respective categories being at least three orders of magnitudes larger than those present in other compartments. It may be possible that the SAEM algorithm in Monolix also struggles with particularly large parameters on certain datasets however this cannot be proven from the results so far.

Moving forward, we reduce the model by eliminating the C and F compartments completely. As mentioned earlier, our justification is that they are de–coupled from the dynamics of B, A, and I, which are more interested in given that they are what drive the antibody dynamics. We conclude that the C and F compartments cannot have their parameter identifiability conclusively analysed with the given parameter distributions and NLME methods. For future work, we suggest further studies into the effectiveness of estimating cytotoxic T–cell and IFN–γ from limited sets of laboratory data.

2.6 Model reduction and identifiability analysis using the simplex algorithm

For the remainder of this study, we consider a reduced version of the full LNP model with $\gamma_L = 0$ and the C and F compartments eliminated. This model is referred to as the *LNP Model* and is given by the following ODEs

$$\frac{dL}{dt} = -\mu_{LV} L \tag{2.20a}$$

$$\frac{dV}{dt} = \mu_{LV} L - \gamma_V V \tag{2.20b}$$

$$\frac{dT}{dt} = \mu_{VT} V - \gamma_T T \tag{2.20c}$$

$$\frac{dB}{dt} = \mu_{TB} T + \beta_{BI} \left(\frac{I}{I + s_I} \right) B - \gamma_B B \tag{2.20d}$$

$$\frac{dA}{dt} = \mu_{BA} B - \gamma_A A \tag{2.20e}$$

$$\frac{dI}{dt} = \mu_{TI} T - \beta_{IB} I B - \gamma_I I \tag{2.20f}$$

The definitions of the compartments, parameters, and units are the same as those outlined in Section 2.1. This is also true for the initial conditions but we re–define them for convenience

$$\begin{aligned} &L(0) = 1, \quad V(0) = 0, \quad T(0) = 0, \\ &B(0) = 0, \quad A(0) = A_0, \quad I(0) = I_0 \end{aligned} \tag{2.21}$$

Since the eliminated compartments C and F do not have any backward feedback, meaning that they do not affect the dynamics of any compartments that are upstream, and are de–coupled from the rest of the model, the differential algebra analysis with all compartments observed will still yield global identifiability. The results are the exact same as those shown in Section 2.2 with the C and F parameters simply not being present. Therefore, we assume for the remainder of this chapter that the LNP model from equations 2.20a – 2.20f is globally identifiable.

The generation of virtual patients in MATLAB is also the same as Section 2.3 but slightly simpler due to the model reduction. Due to the extremely long computation times observed in Sections 2.4 and 2.5 as well as the implied practical identifiability issues in the removed C and F compartments despite the presence of a highly dense dataset, we reduce the number of observed time points substantially. We are assuming here that as long as the distribution of time points is sufficiently dense enough to capture the full peaks (including growth and decay) of each compartment's solution, then any remaining issues with practical identifiability stem from the distributions of the parameters themselves and the fitting algorithms being less suited to the given datasets. Given this, the new timespan $\mathbf{t}$ consists of 101 time points with the first point being at $t = 0$ and the remaining 100 points logarithmically spaced from $t = 0.001$ to $t = 365$ days. The choice of a logarithmic timescale here stems from the fact that the peaks of all compartments almost always occur within one month of the vaccine administration.

One change to the generation of virtual patient parameter vectors $\mathbf{p}_i$ is the removal of variability on the interleukin saturation which is instead fixed to $s_I = 1000$ across all patients. Korosec et al. [27] fixed s_I as such in their study based on the previous work in Farhang–Sardroodi et al. [26] as well as the maximum interleukin values in the clinical datasets from Bergamaschi et al. [30]. While the differential algebra analysis from DAISY and the individual parameter fits of Section 2.5 indicate that s_I is identifiable, we choose to keep s_I as a fixed parameter as it was in Korosec et al. since no information about an appropriate distribution for the parameter is available from the results of fitting the laboratory data. We discuss the case where s_I is unfixed in Section 3.4.

Given all of the above, we generate fifty virtual patients which we will refer to as P_1, ..., P_{50}. The exact parameter values for each patient are shown in Table 2.22. In all such tables, the patient IDs refer to the index i of a given virtual patient P_i. We now discuss fitting parameters from the LNP model in Monolix. As outlined in Section 1.3, the default fitting algorithm used in Monolix is SAEM for which variability is added to parameters via the random effects. In the previous experiments these effects were disabled for any parameter that was fixed. Given that our datasets so far have only ever represented a population of size per fit of $N = 1$, being the simulated data of only a single patient, we consider the case where parameters are fitted without such variability to observe whether or not the errors are improved. We recall that Monolix offers multiple options for dealing with designated non–variable parameters. These include maintaining no variability meaning no random effects (default), adding decreasing variability via random effects, and adding variability via random effects only in the exploratory phase. In this section, we choose to maintain no variability for all parameters regardless of whether or not they are fixed by completely disabling the random effects. This switches the estimation method from SAEM to the Nelder–Mead simplex algorithm.

The auto–initialisation (auto–init) procedure in Monolix is used here and is maintained for the remainder of this study. As outlined in Section 1.3, the auto–init computation uses a proprietary optimisation method on pooled data without inter–individual variability [64]. The algorithm can be run multiple times iteratively with each iterate using the previously calculated initial parameter values. For our implementation of the procedure into the experiment, we start off with all non–fixed initial parameter values set to the expected value (midpoint) of their respective uniform

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	3.47	12.1	0.07	0.54	12.04	0.12	0.097	0.59	0.044	0.053	3.22	0.0032	1000	346.47	163.63
2	4.16	15.81	0.13	0.1	8.8	0.058	0.06	0.024	0.049	0.026	4.61	0.0035	1000	504.78	86.65
3	4.22	14.72	0.068	0.51	27.68	0.13	0.086	0.53	0.05	0.023	2.98	0.0017	1000	52.57	62.44
4	2.23	24.64	0.14	0.12	9.23	0.1	0.12	0.51	0.05	0.041	2.62	0.0017	1000	378.3	154.88
5	1.28	23.34	0.1	0.54	3.29	0.084	0.047	0.024	0.039	0.037	3.77	0.0018	1000	70.21	242.7
6	3.56	8.06	0.13	0.56	14.27	0.075	0.048	0.094	0.045	0.054	3.81	0.0013	1000	182.72	12.36
7	3.56	16.58	0.1	0.32	21.23	0.07	0.067	0.17	0.044	0.029	2.7	0.0018	1000	480.53	193.78
8	0.6	17.93	0.12	0.074	26.29	0.11	0.056	0.33	0.052	0.054	1.99	0.0022	1000	248.39	46.96
9	0.98	20.83	0.11	0.17	11.56	0.087	0.11	0.13	0.04	0.021	2.62	0.0033	1000	149.56	129.33
10	0.26	5.77	0.076	0.99	25.66	0.11	0.11	0.49	0.039	0.039	4.81	0.0023	1000	349.58	198.67
11	0.27	28.18	0.14	0.2	8.96	0.14	0.056	0.31	0.051	0.05	2.63	0.0031	1000	259.96	231.78
12	2.62	24.8	0.13	0.37	27.56	0.08	0.077	0.54	0.047	0.039	3.57	0.0028	1000	331.07	172.34
13	0.7	0.8	0.13	0.62	5.86	0.11	0.12	0.5	0.041	0.04	2.96	0.0016	1000	157.19	135.62
14	0.55	15.19	0.13	0.3	25.36	0.054	0.1	0.55	0.05	0.052	2.44	0.0013	1000	384.12	105.8
15	3.01	25.32	0.1	0.6	23.59	0.12	0.12	0.28	0.038	0.04	2.34	0.0038	1000	106.6	224.14
16	2.34	5.89	0.092	0.18	18.9	0.13	0.1	0.13	0.04	0.032	3.24	0.0031	1000	255.39	144.71
17	3.66	2.69	0.1	0.13	21.2	0.065	0.079	0.59	0.052	0.051	3.59	0.0013	1000	23.15	114.22
18	2.21	29.35	0.12	0.97	23.34	0.075	0.052	0.25	0.036	0.047	3.23	0.0022	1000	99.1	253.62
19	4.06	14.42	0.073	0.096	11.03	0.13	0.088	0.18	0.049	0.039	2.94	0.0033	1000	100.21	99.8
20	4.08	18.95	0.084	0.85	24.13	0.08	0.077	0.55	0.049	0.038	4.69	0.0031	1000	175.44	238.54
21	0.72	16	0.072	0.15	10.39	0.1	0.1	0.095	0.047	0.031	3.96	0.0034	1000	152.22	108.95
22	3.61	16.48	0.067	0.65	22.13	0.087	0.043	0.54	0.046	0.046	2.33	0.0033	1000	182.92	44.15
23	1.74	18.46	0.097	0.69	18.91	0.086	0.12	0.07	0.047	0.051	2.72	0.0023	1000	516.22	147.89
24	1.55	2.56	0.12	0.94	9.32	0.052	0.062	0.55	0.045	0.045	3.66	0.0028	1000	21.79	214.35
25	2.54	1.47	0.074	0.23	9.36	0.098	0.078	0.31	0.049	0.034	3.39	0.0027	1000	327.88	242.14
26	2.94	6.94	0.1	0.17	10.22	0.13	0.049	0.085	0.037	0.04	3.67	0.002	1000	517.14	153.99
27	2.68	11.66	0.13	0.34	11.98	0.12	0.083	0.57	0.049	0.035	4.73	0.0014	1000	170.93	170.42
28	3.18	7.47	0.11	0.92	23.17	0.13	0.045	0.26	0.044	0.02	2.2	0.0025	1000	463.53	57.44
29	3.02	1.17	0.087	0.54	26.24	0.11	0.11	0.56	0.038	0.037	4	0.0023	1000	318.92	40.82
30	1.35	27.04	0.097	0.83	1.58	0.13	0.04	0.31	0.048	0.022	4.07	0.0036	1000	3.65	243.99
31	1.51	20.88	0.14	0.97	24.05	0.1	0.083	0.15	0.04	0.031	4.69	0.0023	1000	156.53	72.49
32	3.19	17.84	0.14	0.67	22.47	0.098	0.046	0.18	0.047	0.038	3.2	0.0038	1000	176.91	64.92
33	3.61	27.96	0.087	0.48	23.86	0.1	0.093	0.32	0.04	0.037	3.8	0.0024	1000	249.03	230.15
34	1.99	1.55	0.073	0.18	22.11	0.094	0.047	0.43	0.04	0.032	4.17	0.0031	1000	149.3	108.12
35	3.15	14.33	0.071	0.65	20.64	0.12	0.047	0.31	0.051	0.031	4.7	0.0038	1000	22.39	190.54
36	0.38	24.82	0.066	0.77	8.02	0.11	0.042	0.47	0.038	0.049	4.7	0.0033	1000	187.31	42.01
37	2.48	7.36	0.14	0.26	6.24	0.073	0.054	0.091	0.044	0.033	2.38	0.0022	1000	40.73	8.93
38	3.98	14.3	0.078	0.73	1.5	0.064	0.061	0.38	0.046	0.049	4.49	0.0036	1000	382.76	68.65
39	2.53	20.08	0.13	0.39	24.37	0.084	0.047	0.35	0.041	0.026	2.58	0.0028	1000	197.23	105.79
40	2.53	24.06	0.085	0.52	10.38	0.14	0.06	0.56	0.038	0.045	3.14	0.0037	1000	500.85	123.59
41	0.58	25.33	0.089	0.23	8.84	0.072	0.052	0.043	0.042	0.035	4.01	0.0037	1000	181.42	104.18
42	0.7	22.36	0.065	0.18	18.83	0.078	0.095	0.57	0.046	0.028	2.31	0.0031	1000	430.16	177.24
43	1.94	12.19	0.13	0.18	12	0.053	0.11	0.4	0.047	0.035	3.95	0.003	1000	322.21	244.44
44	1.49	0.96	0.099	0.4	7.37	0.1	0.077	0.34	0.036	0.025	2.51	0.0035	1000	435.11	150.42
45	1.95	23.99	0.11	0.49	18.93	0.14	0.076	0.3	0.04	0.032	4.23	0.0012	1000	526.65	168.47
46	0.55	11.96	0.1	0.96	23.7	0.07	0.094	0.028	0.048	0.025	3.68	0.0031	1000	411.85	7.24
47	1.83	10.68	0.12	0.64	5.98	0.12	0.091	0.13	0.039	0.042	3.9	0.0013	1000	170.56	25.04
48	3.55	7.77	0.098	0.11	5.58	0.058	0.052	0.26	0.039	0.033	4.62	0.0038	1000	199.66	27.38
49	1.01	26.71	0.14	0.077	14.6	0.12	0.12	0.23	0.043	0.029	4.01	0.0036	1000	14.35	142.17
50	2.07	12.6	0.11	0.82	25.61	0.077	0.12	0.18	0.036	0.036	3.17	0.0034	1000	312.05	135.44

Table 2.22: Parameter values of virtual patients P_1, ..., P_{50} generated in MATLAB for the LNP model. Note that $s_I = 1000$ is fixed for all patients.

distributions. We then run the auto–init procedure multiple times until the initial parameter estimates have stabilised to a new value. We did not utilise a strict convergence criteria for when to stop iteratively applying auto–init. However for almost all virtual patients, the initial parameter values would stop changing after five iterations. One final change to the experimental setup in Monolix is the fixing of the variable non–zero initial values of antibody (A_0) and interleukin (I_0) to their exact values generated in MATLAB. In a similar vein to s_I, we hope that by reducing the number of simultaneously fitted parameters, the errors in the parameter fits are reduced. We justify this under the assumption that the initial concentrations of antibody and interleukin are measured at the time of vaccine administration. We now begin testing on the LNP model with a control case fitting with the simplex algorithm where no additional parameters are fixed. The experimental procedure is the following

1. Load the solution dataset U_i (note that $s_I = 1000$ is now fixed for all patients)
2. Set all compartments as observations
3. Load the LNP model equations
4. Fix s_I, A_0, and I_0 with their initial values fixed to their exact values from MATLAB with random effects disabled
5. Set all other parameters as variable and to be fitted under the maximum likelihood estimation with random effects disabled and an initial value equal to the expected value of the respective uniform distribution
6. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space (5 iterations in most cases)
7. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
8. Run the Nelder–Mead (simplex) algorithm which outputs the population parameter fits, which are equivalent to the individual patient fits

The results are shown in Table 2.23.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	1.15%	9.17%	1.43%	1.85%	3.49%	8.33%	8.25%	3.39%	2.27%	5.66%	7.14%	0.00%
2	0.48%	0.32%	0.00%	0.00%	0.68%	3.45%	3.33%	0.00%	0.00%	15.38%	0.00%	2.86%
3	1.18%	9.58%	1.47%	0.00%	4.01%	0.00%	0.00%	1.89%	0.00%	4.35%	1.68%	11.76%
4	1.35%	10.39%	7.14%	8.33%	15.93%	3.00%	0.00%	3.92%	2.00%	12.20%	0.38%	23.53%
5	0.78%	0.47%	5.00%	9.26%	3.65%	1.19%	0.00%	0.00%	5.13%	29.73%	0.00%	11.11%
6	0.84%	0.87%	7.69%	12.50%	1.61%	0.00%	0.00%	0.00%	0.00%	11.11%	0.52%	15.38%
7	0.28%	1.69%	0.00%	3.13%	1.84%	12.86%	16.42%	0.00%	0.00%	10.34%	0.37%	0.00%
8	0.00%	0.06%	8.33%	14.86%	0.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.01%	13.64%
9	1.02%	5.66%	0.00%	0.00%	0.00%	2.30%	9.09%	0.00%	0.00%	4.76%	3.44%	3.03%
10	0.00%	0.52%	1136.84%	115.15%	100.00%	0.00%	0.00%	18.37%	12.82%	28.21%	35.76%	100.00%
11	3.70%	3.30%	0.00%	5.00%	11.72%	0.00%	0.00%	0.00%	1.96%	10.00%	2.66%	22.58%
12	4.58%	5.40%	7.69%	2.70%	2.79%	6.25%	2.60%	3.70%	0.00%	2.56%	5.60%	0.00%
13	0.00%	23.75%	7.69%	53.23%	36.69%	36.36%	29.17%	12.00%	2.44%	5.00%	14.19%	75.00%
14	14.55%	7.57%	1000.00%	473.33%	59.54%	44.44%	43.00%	56.36%	400.00%	100.00%	88.11%	98.23%
15	3.65%	8.97%	0.00%	0.00%	8.18%	8.33%	16.67%	0.00%	2.63%	7.50%	5.56%	2.63%
16	93.59%	1759170.80%	31.52%	11.11%	2.86%	2720492.31%	18.00%	7.69%	0.00%	146.88%	13.27%	6.45%
17	90.44%	3.72%	790.00%	23.08%	100.00%	16.92%	8.86%	20.34%	30.77%	68.63%	221.17%	100.00%
18	97.74%	2624.63%	33.33%	24.74%	14.31%	406.67%	284.62%	8.00%	5.56%	14.89%	6.81%	0.00%
19	1.48%	11.65%	2.74%	14.58%	4.35%	0.00%	1.14%	0.00%	0.00%	10.26%	4.42%	12.12%
20	0.00%	10.61%	1.19%	35.29%	153.83%	125.00%	45.45%	0.00%	4.08%	1557.89%	4.48%	112.90%
21	220.83%	23.25%	36.11%	0.00%	98.08%	20.00%	40.00%	110.53%	6.38%	99.65%	120.71%	97.29%
22	0.83%	0.79%	2.99%	6.15%	1.13%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	6.06%
23	0.00%	12.30%	13.40%	5.80%	15.18%	39.53%	27.50%	0.00%	0.00%	15.69%	0.00%	8.70%
24	1.29%	2.73%	166.67%	167.02%	99.92%	26.92%	19.35%	23.64%	2.22%	75.56%	18.58%	100.00%
25	10.63%	10.88%	1.35%	21.74%	19.12%	12.24%	5.13%	3.23%	0.00%	11.76%	1.18%	22.22%
26	0.00%	0.14%	0.00%	5.88%	5.68%	0.00%	0.00%	0.00%	0.00%	30.00%	1.36%	10.00%
27	2.24%	7.89%	7.69%	8.82%	8.51%	8.33%	4.82%	8.77%	2.04%	5.71%	12.05%	7.14%
28	11.95%	21.02%	10.91%	354.35%	63.53%	7.69%	22.22%	23.08%	195.45%	10.00%	14.55%	148.00%
29	1.32%	17.09%	0.00%	1.85%	11.24%	18.18%	14.55%	3.57%	0.00%	2.70%	11.00%	4.35%
30	12.59%	4.59%	98.25%	0.00%	60109.49%	33.08%	7.50%	93070.97%	4.17%	9900.00%	47440.54%	56011.11%
31	1.32%	2.01%	0.00%	2.06%	1.00%	0.00%	3.61%	0.00%	0.00%	9.68%	5.54%	8.70%
32	98.53%	44132.40%	99.86%	14.93%	63.33%	33787.76%	28.26%	69783.33%	4.26%	75.79%	20390.63%	89.74%
33	30.75%	46.60%	26.44%	50.00%	73.55%	8.00%	29.03%	3.13%	0.00%	59.46%	21.84%	79.58%
34	10.05%	47.74%	99.85%	11.11%	2056.35%	39.36%	29.79%	5665.12%	10.00%	3837.50%	2782.01%	706.45%
35	98.44%	5183.81%	11.27%	29.23%	52.62%	58.33%	3006.38%	6.45%	5.88%	94.19%	12.77%	55.26%
36	44.74%	71.96%	20839.39%	80.52%	60.97%	63.64%	0.00%	179672.34%	2.63%	93.88%	60324.04%	77.88%
37	1.21%	1.90%	0.00%	3.85%	3.21%	1.37%	1.85%	1.10%	0.00%	18.18%	0.42%	4.55%
38	3.27%	5.45%	1.28%	21.92%	41.33%	14.06%	8.20%	36.84%	4.35%	26.53%	43.21%	47.22%
39	0.00%	0.05%	0.00%	10.26%	0.98%	0.00%	0.00%	0.00%	0.00%	0.00%	1.94%	10.71%
40	1.98%	0.58%	3.53%	5.77%	3.47%	7.14%	1.67%	0.00%	0.00%	4.44%	0.64%	8.11%
41	0.00%	0.24%	7.87%	13.04%	5.09%	1.39%	0.00%	2.33%	2.38%	57.14%	1.00%	18.92%
42	1.43%	9.30%	4.62%	22.22%	15.08%	14.10%	14.74%	3.51%	2.17%	3.57%	4.33%	22.58%
43	0.00%	1.64%	7.69%	5.56%	0.83%	0.00%	0.00%	5.00%	0.00%	11.43%	2.78%	3.33%
44	5.37%	2.08%	3.03%	10.00%	3.12%	20.00%	11.69%	2.94%	2.78%	4.00%	3.98%	2.86%
45	0.51%	1.25%	9.09%	40.82%	18.81%	42.86%	21.05%	3.33%	2.50%	337.50%	3.78%	41.67%
46	0.00%	6.86%	2.00%	7.29%	8.35%	0.00%	2.13%	0.00%	2.08%	36.00%	0.27%	9.68%
47	0.55%	8.90%	0.00%	3.13%	2.17%	0.00%	3.30%	0.00%	0.00%	2.38%	8.72%	15.38%
48	1.13%	1.29%	2.04%	9.09%	2.15%	1.72%	1.92%	0.00%	0.00%	3.03%	0.22%	5.26%
49	93.47%	796.86%	85.00%	1341.56%	384.73%	616.67%	200.00%	8.70%	2.33%	2210.34%	11.97%	2400.00%
50	12.08%	36.19%	45.45%	3.66%	40.73%	1.30%	0.00%	0.00%	0.00%	5.56%	5.99%	2.94%
Arithmetic Mean	19.67%	36247.34%	492.56%	61.32%	1275.91%	55120.22%	79.83%	6971.47%	14.43%	381.74%	2633.26%	1212.74%
Median	1.33%	7.21%	7.42%	9.63%	9.88%	8.17%	6.31%	3.18%	2.02%	11.98%	5.01%	12.88%

Table 2.23: Percentage errors from fitting all parameters simultaneously using the simplex algorithm on virtual patients P_1, ..., P_{50} generated from the LNP model. Also shown are the mean and median errors for each parameter. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Errors exceeding 100% all have the same shade with a new shade at 1000% to indicate extremely divergent fits. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Similar to the previous case where all parameters were simultaneously fitted in Table 2.11 they are once again not fitted consistently well; another discrepancy with the global identifiability prediction from DAISY. Still, with a much larger population size of $N = 50$ compared to the previous $N = 5$, we can say that certain parameters are fitted better than others. For example, μ_{LV} and γ_A are clearly fitted much better than μ_{TB} and γ_I. We still however observe completely divergent parameter fits with percentage errors in the thousands or millions. One pattern that does appear to emerge is that the errors in fitting are consistent across a given patient rather than across a given parameter. For example, compare patients 1–9 which have consistently low errors versus patients 32–36 which have consistently higher errors. This suggests that the model may be sensitive to particular parameter values that make them harder to identify.

We consider a way to improve the fits and also isolate particular parameters that may not be as well suited for the given fitting procedures. One way we can do this is to progressively fix parameters one by one while trying to minimise the total number that are fixed. Ideally, we expect fixing parameters to reduce the errors by giving Monolix a more complete picture of the dynamics on specific timescales. For example, Moyles et al. [35] showed that the priming of vaccinated cells (V) from LNP (L) occurs on a timescale of $t \sim \mathcal{O}(\mu_{LV}^{-1})$. So if μ_{LV} is known, then the timescale can be calculated exactly. We therefore repeat the experiment fitting virtual patients P_1, ..., P_{50} with various sets of parameters fixed to their exact values for each patient. We inform our decisions on which particular parameters to fix based on a combination of the mean and median errors seen in Table 2.23 while trying to minimise the number of fixed parameters as well as, to a lesser extent, the number of compartments that include them. The experimental procedure is similar to the control with the additional step of fixing a set of parameters in addition to s_I, A_0, and I_0 which as mentioned are fixed in all cases

1. Load the solution dataset U_i (note that $s_I = 1000$ is no longer variable and is fixed for all patients)
2. Set all compartments as observations
3. Load the LNP model equations
4. Fix s_I, A_0, and I_0 with their initial values fixed to their exact values from MATLAB with random effects disabled
5. Fix additional parameters given the previous mean errors while trying to minimise the total number that are fixed

6. Set all other parameters as variable and to be fitted under the maximum likelihood estimation with random effects disabled and an initial value equal to the expected value of the respective uniform distribution

7. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space (5 iterations in most cases)

8. Set posterior distributions for data on all compartments to normal distributions (Monolix default)

9. Run the Nelder–Mead (simplex) algorithm which outputs the population parameter fits, which are equivalent to the individual patient fits

We fit the virtual patient population under four different fixed parameter sets. The results are summarised in Table 2.24 and yield two main takeaways. Firstly, fixing at least one of the autocatalytic rates β_{BI} or β_{IB} produced substantially improved errors. We suspect that this may be due to the non–linear terms introducing more coupling into the model and making it harder to separate the individual parameters. Therefore, fixing one of the non–linear rates is beneficial. Secondly, fixing β_{BI} in the third row as an additional fixed parameter compared to the second row caused an increase in the mean error despite there being less variable parameters. This may be due to the propagation of errors from parameters that get fixed to parameters that are not however it is difficult to say this conclusively from the results thus far.

Fixed Parameters	Mean Error	Results Table
μ_{VT}, γ_V	10.59%	2.25
μ_{LV}, γ_V, β_{IB}	3.24%	2.26
μ_{LV}, γ_V, β_{BI}, β_{IB}	4.67%	2.27
μ_{VT}, γ_V, β_{BI}, β_{IB}	2.55%	2.28

Table 2.24: Summary of parameter fits attained using the simplex algorithm in Monolix to fit virtual patients P_1, ..., P_{50} with various fixed parameter sets. The mean errors are calculated by taking the mean of all individual parameter errors across all patients. The specified results tables showcase the full results.

That being said, given the fixed parameter set that resulted in the lowest mean error, we can conclude that prior individual patient knowledge of non–linear dynamics (β_{BI}, β_{IB}) as well as the efficiency of T–cell priming via vaccinated cells (γ_V, μ_{VT}) can be crucial insights for improving individual parameter fits for a given patient. This itself is not very surprising. Non–linear systems

cannot be solved analytically and the T-cell dynamics are what the rest of the immune response stems from.

Part of the reason why we disabled the random effects was that initial tests with the SAEM algorithm on a single set of individual patient data showed higher errors than the simplex algorithm with random effects disabled. Of course, recalling equation 1.8, one of the main steps of SAEM is the iterative generation of individual parameter vectors $\mathbf{\Psi}_i$ using the conditional distribution given the observed individual dataset U_i and population parameters $\mathbf{\Theta}_k$ at the previous iteration k. Our previous testing with SAEM only used a single patient dataset in each fit where the population size per fit is $N = 1$. Since the true strength of SAEM comes from fitting using data from a larger population, in the next section, we consider fitting the parameters with SAEM once again but with an increased population size.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	17.87%	0.00%	7.14%	16.67%	2.41%	0.00%	1.03%	1.69%	2.27%	3.77%	0.93%	12.50%
2	0.00%	0.00%	51.54%	110.00%	2.61%	0.00%	0.00%	0.00%	2.04%	99.04%	0.43%	102.86%
3	2.13%	0.00%	4.41%	1.96%	8.42%	0.00%	1.16%	1.89%	2.00%	4.35%	2.01%	0.00%
4	5.83%	0.00%	0.00%	8.33%	11.48%	0.00%	0.00%	3.92%	2.00%	4.88%	8.40%	11.76%
5	0.00%	0.00%	4.00%	9.26%	4.56%	0.00%	0.00%	0.00%	2.56%	59.46%	0.53%	11.11%
6	0.00%	0.00%	7.69%	14.29%	0.07%	0.00%	0.00%	0.00%	0.00%	3.70%	0.52%	15.38%
7	0.00%	0.00%	2.00%	3.13%	0.28%	0.00%	0.00%	0.00%	0.00%	3.45%	0.37%	5.56%
8	0.00%	0.00%	8.33%	27.03%	0.80%	0.00%	0.00%	3.03%	0.00%	0.00%	3.02%	27.27%
9	12.24%	0.00%	0.00%	0.00%	4.50%	0.00%	9.09%	0.00%	0.00%	0.00%	5.34%	3.03%
10	11.54%	0.00%	1.32%	6.06%	2.84%	0.00%	0.00%	0.00%	2.56%	0.00%	7.69%	17.39%
11	0.00%	0.00%	14.29%	20.00%	4.35%	0.00%	0.00%	0.00%	0.00%	2.00%	1.52%	22.58%
12	0.00%	0.00%	0.00%	5.41%	19.12%	0.00%	1.30%	5.56%	0.00%	0.00%	8.12%	17.86%
13	20.00%	0.00%	7.69%	19.35%	3.41%	0.00%	16.67%	6.00%	0.00%	0.00%	4.05%	25.00%
14	0.00%	0.00%	0.00%	16.67%	9.58%	0.00%	0.00%	3.64%	0.00%	7.69%	0.82%	23.08%
15	21.59%	0.00%	1.00%	10.00%	1.87%	0.00%	0.00%	3.57%	0.00%	7.50%	2.56%	10.53%
16	9.83%	0.00%	1.09%	0.00%	8.15%	0.00%	2.00%	0.00%	2.50%	12.50%	6.17%	6.45%
17	11.48%	0.00%	4.00%	7.69%	1.51%	0.00%	1.27%	5.08%	1.92%	1.96%	0.56%	15.38%
18	15.38%	0.00%	25.00%	19.59%	53.34%	0.00%	0.00%	12.00%	0.00%	94.26%	29.10%	54.55%
19	23.89%	0.00%	2.74%	4.17%	3.90%	0.00%	2.27%	0.00%	2.04%	12.82%	4.42%	3.03%
20	19.36%	0.00%	0.00%	18.82%	40.12%	0.00%	1.30%	20.00%	4.08%	21.05%	27.93%	41.94%
21	19.44%	0.00%	13.89%	0.00%	12.70%	0.00%	0.00%	1.05%	0.00%	9.68%	0.76%	11.76%
22	0.00%	0.00%	2.99%	6.15%	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	6.06%
23	0.00%	0.00%	3.09%	2.90%	2.22%	0.00%	0.00%	0.00%	0.00%	25.49%	0.00%	4.35%
24	0.00%	0.00%	8.33%	9.57%	2.68%	0.00%	1.61%	3.64%	0.00%	11.11%	1.09%	10.71%
25	1.18%	0.00%	12.16%	17.39%	1.28%	0.00%	1.28%	3.23%	0.00%	5.88%	1.77%	14.81%
26	0.00%	0.00%	3.00%	5.88%	5.97%	0.00%	0.00%	0.00%	0.00%	22.50%	1.09%	10.00%
27	13.81%	0.00%	0.00%	2.94%	0.92%	0.00%	1.20%	1.75%	0.00%	2.86%	2.75%	14.29%
28	0.00%	0.00%	0.00%	6.52%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.45%	8.00%
29	30.46%	0.00%	3.45%	5.56%	3.66%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.70%
30	0.00%	0.00%	19.59%	26.51%	98.23%	0.00%	0.00%	73.87%	10.42%	100.00%	1.23%	75.00%
31	9.27%	0.00%	7.14%	1.03%	4.74%	0.00%	2.41%	0.00%	0.00%	0.00%	7.25%	4.35%
32	0.00%	0.00%	0.00%	5.97%	1.07%	0.00%	0.00%	0.00%	0.00%	2.63%	0.94%	7.89%
33	12.74%	0.00%	3.45%	0.00%	5.11%	0.00%	2.15%	0.00%	2.50%	2.70%	2.11%	0.00%
34	0.00%	0.00%	53.42%	5.56%	12.98%	0.00%	0.00%	32.56%	2.50%	0.00%	27.10%	9.68%
35	0.32%	0.00%	2.82%	4.62%	0.39%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.26%
36	0.00%	0.00%	4.55%	10.39%	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.64%	12.12%
37	0.00%	0.00%	0.00%	7.69%	1.28%	0.00%	0.00%	1.10%	0.00%	6.06%	1.68%	9.09%
38	0.75%	0.00%	3.85%	5.48%	3.33%	0.00%	0.00%	2.63%	0.00%	4.08%	1.56%	8.33%
39	0.00%	0.00%	0.00%	12.82%	0.66%	0.00%	0.00%	0.00%	0.00%	0.00%	1.55%	10.71%
40	9.09%	0.00%	1.18%	9.62%	3.76%	0.00%	0.00%	1.79%	0.00%	0.00%	0.96%	13.51%
41	0.00%	0.00%	5.62%	4.35%	5.77%	0.00%	0.00%	0.00%	0.00%	71.43%	0.25%	8.11%
42	14.29%	0.00%	3.08%	0.00%	13.70%	0.00%	3.16%	1.75%	0.00%	0.00%	8.23%	9.68%
43	0.52%	0.00%	7.69%	27.78%	91.00%	0.00%	0.00%	70.00%	2.13%	73.71%	102.28%	91.33%
44	69.13%	0.00%	50.51%	35.00%	40.57%	0.00%	10.39%	8.82%	5.56%	40.00%	4.78%	137.14%
45	6.67%	0.00%	14.55%	77.55%	99.10%	0.00%	5.26%	276.67%	5.00%	97.59%	366.19%	99.60%
46	0.00%	0.00%	5.00%	20.83%	1.14%	0.00%	0.00%	0.00%	4.17%	48.00%	1.36%	16.13%
47	16.39%	0.00%	0.00%	1.56%	1.17%	0.00%	3.30%	0.00%	0.00%	7.14%	6.41%	7.69%
48	0.00%	0.00%	3.06%	9.09%	0.90%	0.00%	0.00%	0.00%	0.00%	3.03%	0.22%	7.89%
49	14.85%	0.00%	0.00%	1.30%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	5.24%	11.11%
50	13.04%	0.00%	0.00%	2.44%	0.51%	0.00%	0.00%	0.00%	0.00%	2.78%	1.89%	5.88%
Arithmetic Mean	7.96%	0.00%	7.65%	13.11%	12.22%	0.00%	1.36%	11.13%	1.15%	17.80%	13.53%	21.64%
Median	0.97%	0.00%	3.27%	7.11%	3.37%	0.00%	0.00%	0.00%	0.00%	3.74%	1.62%	10.91%

Table 2.25: Percentage errors from fitting parameters simultaneously using the simplex algorithm on virtual patients P_1, ..., P_{50} generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT} and γ_V are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	0.00%	5.21%	1.43%	1.85%	0.17%	0.00%	2.06%	6.78%	0.00%	1.89%	5.59%	0.00%
2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%
3	0.00%	8.63%	10.29%	7.84%	1.48%	0.00%	0.00%	5.66%	2.00%	0.00%	4.70%	0.00%
4	0.00%	10.47%	0.00%	8.33%	1.19%	0.00%	0.00%	1.96%	2.00%	2.44%	5.73%	0.00%
5	0.00%	0.04%	2.00%	1.85%	1.22%	0.00%	0.00%	0.00%	0.00%	2.70%	0.27%	0.00%
6	0.00%	0.37%	0.00%	0.00%	0.91%	0.00%	0.00%	0.00%	0.00%	7.41%	0.00%	0.00%
7	0.00%	5.19%	0.00%	3.13%	0.42%	0.00%	2.99%	0.00%	0.00%	13.79%	1.85%	0.00%
8	0.00%	0.17%	0.00%	0.00%	0.34%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%	0.00%
9	0.00%	9.17%	0.00%	0.00%	0.61%	0.00%	0.00%	0.00%	0.00%	4.76%	7.25%	0.00%
10	0.00%	9.71%	2.63%	1.01%	0.08%	0.00%	0.00%	4.08%	2.56%	5.13%	9.56%	0.00%
11	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.00%	0.00%	2.00%	0.76%	0.00%
12	0.00%	35.24%	23.08%	13.51%	29.57%	0.00%	1.30%	1.85%	0.00%	7.69%	5.88%	0.00%
13	0.00%	27.50%	0.00%	46.77%	23.89%	0.00%	0.00%	16.00%	2.44%	2.50%	25.34%	0.00%
14	0.00%	0.00%	7.69%	6.67%	6.19%	0.00%	0.00%	9.09%	0.00%	5.77%	0.41%	0.00%
15	0.00%	11.02%	0.00%	1.67%	0.13%	0.00%	0.00%	0.00%	2.63%	2.50%	5.56%	0.00%
16	0.00%	10.87%	2.17%	0.00%	0.85%	0.00%	0.00%	0.00%	0.00%	18.75%	5.86%	0.00%
17	0.00%	5.20%	0.00%	7.69%	1.79%	0.00%	2.53%	3.39%	0.00%	3.92%	3.90%	0.00%
18	0.00%	0.03%	0.00%	0.00%	0.60%	0.00%	0.00%	0.00%	0.00%	6.38%	0.00%	0.00%
19	0.00%	9.99%	2.74%	1.04%	1.00%	0.00%	0.00%	0.00%	2.04%	5.13%	10.88%	0.00%
20	0.00%	15.94%	2.38%	2.35%	17.03%	0.00%	0.00%	3.64%	0.00%	0.00%	3.20%	0.00%
21	0.00%	11.94%	9.72%	0.00%	15.11%	0.00%	0.00%	1.05%	0.00%	25.81%	0.76%	0.00%
22	0.00%	0.00%	1.49%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	0.00%
23	0.00%	8.45%	13.40%	1.45%	10.15%	0.00%	0.00%	0.00%	0.00%	9.80%	0.37%	0.00%
24	0.00%	3.13%	0.00%	4.26%	0.43%	0.00%	1.61%	7.27%	0.00%	4.44%	5.74%	0.00%
25	0.00%	10.88%	1.35%	4.35%	4.91%	0.00%	2.56%	6.45%	0.00%	0.00%	5.31%	0.00%
26	0.00%	0.29%	0.00%	0.00%	0.98%	0.00%	0.00%	1.18%	0.00%	10.00%	0.27%	0.00%
27	0.00%	18.10%	7.69%	0.00%	10.93%	0.00%	0.00%	0.00%	2.04%	2.86%	6.98%	0.00%
28	0.00%	2.28%	0.00%	1.09%	1.68%	0.00%	0.00%	0.00%	0.00%	0.00%	0.91%	0.00%
29	0.00%	16.24%	10.34%	5.56%	10.79%	0.00%	0.00%	8.93%	0.00%	5.41%	11.00%	0.00%
30	0.00%	0.00%	1.03%	1.20%	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%
31	0.00%	12.21%	0.00%	1.03%	5.82%	0.00%	1.20%	0.00%	0.00%	9.68%	6.61%	0.00%
32	0.00%	0.11%	0.00%	0.00%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
33	0.00%	7.22%	6.90%	0.00%	1.47%	0.00%	1.08%	3.13%	0.00%	5.41%	3.42%	0.00%
34	0.00%	0.00%	4.11%	0.00%	1.18%	0.00%	0.00%	4.65%	0.00%	12.50%	0.48%	0.00%
35	0.00%	0.00%	0.00%	1.54%	1.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	0.00%	0.00%	3.03%	0.00%	0.50%	0.00%	0.00%	2.13%	0.00%	0.00%	1.06%	0.00%
37	0.00%	0.14%	0.00%	3.85%	0.64%	0.00%	0.00%	2.20%	0.00%	12.12%	1.26%	0.00%
38	0.00%	0.14%	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.34%	0.00%
39	0.00%	0.00%	0.00%	0.00%	0.29%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	0.00%	2.54%	8.24%	5.77%	2.60%	0.00%	1.67%	7.14%	0.00%	6.67%	5.10%	0.00%
41	0.00%	0.00%	1.12%	0.00%	0.23%	0.00%	0.00%	0.00%	0.00%	2.86%	0.25%	0.00%
42	0.00%	11.58%	7.69%	5.56%	3.82%	0.00%	3.16%	3.51%	2.17%	7.14%	4.33%	0.00%
43	0.00%	1.23%	0.00%	5.56%	1.75%	0.00%	0.00%	2.50%	0.00%	0.00%	2.03%	0.00%
44	0.00%	21.88%	1.01%	10.00%	10.18%	0.00%	0.00%	0.00%	2.78%	4.00%	13.15%	0.00%
45	0.00%	10.59%	9.09%	4.08%	0.00%	0.00%	1.32%	0.00%	0.00%	6.25%	6.62%	0.00%
46	0.00%	0.00%	6.00%	0.00%	0.17%	0.00%	0.00%	0.00%	0.00%	20.00%	0.82%	0.00%
47	0.00%	8.90%	0.00%	1.56%	1.00%	0.00%	1.10%	0.00%	0.00%	2.38%	8.46%	0.00%
48	0.00%	0.77%	1.02%	0.00%	0.36%	0.00%	0.00%	3.85%	0.00%	3.03%	2.16%	0.00%
49	0.00%	10.78%	7.14%	1.30%	4.25%	0.00%	0.00%	0.00%	0.00%	10.34%	5.24%	0.00%
50	0.00%	12.30%	0.00%	8.54%	3.05%	0.00%	0.00%	0.00%	0.00%	8.33%	7.57%	0.00%
Arithmetic Mean	0.00%	6.61%	3.19%	3.30%	3.65%	0.00%	0.46%	2.17%	0.42%	5.25%	3.91%	0.00%
Median	0.00%	5.20%	1.08%	1.25%	1.00%	0.00%	0.00%	0.00%	0.00%	3.96%	2.68%	0.00%

Table 2.26: Percentage errors from fitting parameters simultaneously using the simplex algorithm on virtual patients P_1, ..., P_{50} generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{LV}, γ_V, and β_{IB} are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	0.00%	12.48%	12.86%	11.11%	5.15%	0.00%	3.09%	5.08%	4.55%	1.89%	0.00%	0.00%
2	0.00%	0.00%	0.00%	0.00%	0.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3	0.00%	5.98%	7.35%	3.92%	2.71%	0.00%	0.00%	1.89%	0.00%	4.35%	0.00%	0.00%
4	0.00%	11.16%	14.29%	8.33%	2.93%	0.00%	0.00%	1.96%	2.00%	4.88%	0.00%	0.00%
5	0.00%	0.00%	0.00%	0.00%	0.91%	0.00%	0.00%	0.00%	0.00%	2.70%	0.00%	0.00%
6	0.00%	0.12%	0.00%	0.00%	0.35%	0.00%	0.00%	0.00%	0.00%	3.70%	0.00%	0.00%
7	0.00%	4.34%	10.00%	3.13%	4.38%	0.00%	0.00%	0.00%	0.00%	3.45%	0.00%	0.00%
8	0.00%	0.11%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
9	0.00%	8.64%	18.18%	0.00%	6.31%	0.00%	0.00%	0.00%	0.00%	4.76%	0.00%	0.00%
10	0.00%	9.36%	15.79%	5.05%	3.59%	0.00%	0.00%	4.08%	0.00%	2.56%	0.00%	0.00%
11	0.00%	0.00%	0.00%	0.00%	0.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12	0.00%	10.60%	15.38%	8.11%	6.24%	0.00%	1.30%	1.85%	0.00%	2.56%	0.00%	0.00%
13	0.00%	20.00%	7.69%	37.10%	20.99%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
14	0.00%	0.07%	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
15	0.00%	4.03%	10.00%	1.67%	0.00%	0.00%	0.00%	3.57%	5.26%	5.00%	0.00%	0.00%
16	0.00%	5.60%	8.70%	0.00%	4.81%	0.00%	2.00%	0.00%	0.00%	6.25%	0.00%	0.00%
17	0.00%	4.46%	10.00%	0.00%	3.44%	0.00%	1.27%	0.00%	0.00%	3.92%	0.00%	0.00%
18	0.00%	0.03%	0.00%	0.00%	0.17%	0.00%	0.00%	0.00%	0.00%	4.26%	0.00%	0.00%
19	0.00%	7.98%	12.33%	1.04%	0.00%	0.00%	1.14%	5.56%	0.00%	2.56%	0.00%	0.00%
20	0.00%	6.81%	13.10%	2.35%	3.65%	0.00%	0.00%	0.00%	0.00%	2.63%	0.00%	0.00%
21	0.00%	36.81%	66.67%	0.00%	57.36%	0.00%	0.00%	0.00%	0.00%	22.58%	0.00%	0.00%
22	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
23	0.00%	14.14%	13.40%	0.00%	17.66%	0.00%	0.00%	0.00%	0.00%	9.80%	0.00%	0.00%
24	0.00%	0.39%	0.00%	0.00%	1.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
25	0.00%	7.48%	12.16%	4.35%	6.84%	0.00%	0.00%	0.00%	2.04%	0.00%	0.00%	0.00%
26	0.00%	0.00%	0.00%	0.00%	0.39%	0.00%	0.00%	0.00%	0.00%	5.00%	0.00%	0.00%
27	0.00%	10.46%	0.00%	5.88%	13.44%	0.00%	1.20%	1.75%	2.04%	5.71%	0.00%	0.00%
28	0.00%	0.27%	0.00%	1.09%	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
29	0.00%	67.52%	164.37%	12.96%	187.77%	0.00%	0.00%	1.79%	0.00%	0.00%	0.00%	0.00%
30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
31	0.00%	26.77%	42.86%	5.15%	24.86%	0.00%	1.20%	6.67%	0.00%	9.68%	0.00%	0.00%
32	0.00%	0.06%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
33	0.00%	3.72%	9.20%	4.17%	1.72%	0.00%	2.15%	6.25%	2.50%	5.41%	0.00%	0.00%
34	0.00%	7.10%	15.07%	0.00%	6.60%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%	0.00%
35	0.00%	0.07%	0.00%	0.00%	0.39%	0.00%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%
36	0.00%	0.00%	1.52%	0.00%	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
37	0.00%	1.63%	0.00%	0.00%	1.76%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	0.00%
38	0.00%	0.00%	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
39	0.00%	0.05%	0.00%	0.00%	0.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	0.00%	5.11%	5.88%	0.00%	5.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
41	0.00%	0.04%	4.49%	0.00%	0.57%	0.00%	0.00%	0.00%	0.00%	20.00%	0.00%	0.00%
42	0.00%	21.91%	32.31%	0.00%	27.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
43	0.00%	0.08%	0.00%	0.00%	0.50%	0.00%	0.00%	0.00%	0.00%	5.71%	0.00%	0.00%
44	0.00%	22.92%	21.21%	20.00%	28.77%	0.00%	2.60%	0.00%	11.11%	0.00%	0.00%	0.00%
45	0.00%	5.59%	9.09%	4.08%	1.58%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
46	0.00%	1.25%	0.00%	5.21%	0.25%	0.00%	1.06%	0.00%	6.25%	0.00%	0.00%	0.00%
47	0.00%	2.06%	8.33%	1.56%	1.67%	0.00%	0.00%	0.00%	0.00%	4.76%	0.00%	0.00%
48	0.00%	0.26%	0.00%	0.00%	0.54%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	0.00%
49	0.00%	7.49%	14.29%	2.60%	5.96%	0.00%	0.00%	0.00%	0.00%	10.34%	0.00%	0.00%
50	0.00%	16.98%	27.27%	2.44%	15.78%	0.00%	0.00%	0.00%	0.00%	2.78%	0.00%	0.00%
Arithmetic Mean	0.00%	7.24%	11.79%	3.04%	9.38%	0.00%	0.35%	0.83%	0.73%	3.41%	0.00%	0.00%
Median	0.00%	4.19%	7.52%	0.00%	1.70%	0.00%	0.00%	0.00%	0.00%	2.67%	0.00%	0.00%

Table 2.27: Percentage errors from fitting parameters simultaneously using the simplex algorithm on virtual patients P_1, ..., P_{50} generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{LV}, γ_V, β_{BI}, and β_{IB} are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	27.95%	0.00%	7.14%	9.26%	3.90%	0.00%	3.09%	1.69%	2.27%	5.66%	0.00%	0.00%
2	1.68%	0.00%	0.00%	0.00%	0.34%	0.00%	0.00%	0.00%	0.00%	7.69%	0.00%	0.00%
3	0.24%	0.00%	7.35%	5.88%	0.29%	0.00%	0.00%	1.89%	2.00%	4.35%	0.00%	0.00%
4	8.07%	0.00%	0.00%	8.33%	3.68%	0.00%	0.00%	1.96%	2.00%	2.44%	0.00%	0.00%
5	0.00%	0.00%	0.00%	1.85%	1.22%	0.00%	0.00%	0.00%	2.56%	29.73%	0.00%	0.00%
6	0.00%	0.00%	0.00%	0.00%	0.21%	0.00%	0.00%	0.00%	0.00%	3.70%	0.00%	0.00%
7	0.00%	0.00%	0.00%	0.00%	0.47%	0.00%	0.00%	0.00%	0.00%	10.34%	0.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
9	10.20%	0.00%	9.09%	0.00%	2.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
10	0.00%	0.00%	18.42%	2.02%	5.22%	0.00%	0.00%	2.04%	0.00%	7.69%	0.00%	0.00%
11	7.41%	0.00%	0.00%	5.00%	1.34%	0.00%	0.00%	0.00%	1.96%	2.00%	0.00%	0.00%
12	6.87%	0.00%	7.69%	2.70%	0.83%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
13	24.29%	0.00%	15.38%	30.65%	10.58%	0.00%	16.67%	0.00%	0.00%	5.00%	0.00%	0.00%
14	0.00%	0.00%	7.69%	3.33%	0.67%	0.00%	0.00%	1.82%	0.00%	0.00%	0.00%	0.00%
15	14.95%	0.00%	10.00%	0.00%	6.57%	0.00%	0.00%	0.00%	2.63%	5.00%	0.00%	0.00%
16	6.84%	0.00%	4.35%	5.56%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
17	4.64%	0.00%	1.00%	0.00%	0.05%	0.00%	0.00%	0.00%	1.92%	3.92%	0.00%	0.00%
18	0.00%	0.00%	0.00%	1.03%	0.17%	0.00%	0.00%	0.00%	0.00%	2.13%	0.00%	0.00%
19	15.27%	0.00%	4.11%	14.58%	6.89%	0.00%	1.14%	0.00%	4.08%	2.56%	0.00%	0.00%
20	13.24%	0.00%	14.29%	4.71%	4.81%	0.00%	2.60%	1.82%	0.00%	2.63%	0.00%	0.00%
21	4.17%	0.00%	11.11%	6.67%	1.25%	0.00%	10.00%	0.00%	0.00%	16.13%	0.00%	0.00%
22	0.00%	0.00%	1.49%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
23	0.00%	0.00%	2.06%	1.45%	2.33%	0.00%	0.00%	0.00%	0.00%	25.49%	0.00%	0.00%
24	0.65%	0.00%	0.00%	0.00%	1.18%	0.00%	0.00%	0.00%	0.00%	2.22%	0.00%	0.00%
25	1.97%	0.00%	5.41%	4.35%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
26	0.00%	0.00%	0.00%	0.00%	1.66%	0.00%	0.00%	1.18%	0.00%	22.50%	0.00%	0.00%
27	10.07%	0.00%	7.69%	2.94%	1.09%	0.00%	1.20%	1.75%	0.00%	0.00%	0.00%	0.00%
28	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
29	29.14%	0.00%	4.60%	1.85%	0.23%	0.00%	9.09%	3.57%	0.00%	2.70%	0.00%	0.00%
30	0.00%	0.00%	0.00%	0.00%	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
31	3.97%	0.00%	7.14%	1.03%	0.46%	0.00%	1.20%	0.00%	0.00%	0.00%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%	0.00%	0.00%
33	4.99%	0.00%	0.00%	0.00%	5.87%	0.00%	1.08%	3.13%	2.50%	10.81%	0.00%	0.00%
34	0.50%	0.00%	4.11%	0.00%	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
35	0.00%	0.00%	0.00%	1.54%	0.39%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	0.00%	0.00%	1.52%	0.00%	0.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
37	1.21%	0.00%	0.00%	0.00%	0.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
38	0.00%	0.00%	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
39	0.00%	0.00%	0.00%	0.00%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	0.79%	0.00%	4.71%	1.92%	1.93%	0.00%	0.00%	0.00%	0.00%	2.22%	0.00%	0.00%
41	0.00%	0.00%	1.12%	0.00%	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
42	0.00%	0.00%	3.08%	11.11%	1.12%	0.00%	3.16%	1.75%	2.17%	0.00%	0.00%	0.00%
43	0.52%	0.00%	0.00%	5.56%	0.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
44	0.67%	0.00%	1.01%	5.00%	1.63%	0.00%	2.60%	2.94%	0.00%	0.00%	0.00%	0.00%
45	1.54%	0.00%	0.00%	6.12%	5.18%	0.00%	1.32%	0.00%	0.00%	9.38%	0.00%	0.00%
46	0.00%	0.00%	2.00%	3.13%	0.17%	0.00%	0.00%	0.00%	4.17%	8.00%	0.00%	0.00%
47	19.67%	0.00%	16.67%	0.00%	3.85%	0.00%	4.40%	0.00%	2.56%	7.14%	0.00%	0.00%
48	0.00%	0.00%	1.02%	0.00%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
49	7.92%	0.00%	7.14%	2.60%	0.89%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
50	16.43%	0.00%	9.09%	2.44%	1.29%	0.00%	8.33%	5.56%	0.00%	5.56%	0.00%	0.00%
Arithmetic Mean	4.68%	0.00%	3.87%	3.06%	1.69%	0.00%	1.17%	0.52%	0.63%	4.16%	0.00%	0.00%
Median	0.66%	0.00%	1.50%	1.49%	0.67%	0.00%	0.00%	0.00%	0.00%	2.17%	0.00%	0.00%
Arithmetic Mean	1.90%	0.00%	3.56%	6.17%	1.21%	0.00%	0.25%	0.17%	0.00%	2.81%	0.00%	0.00%
Median	0.00%	0.00%	0.00%	0.00%	0.65%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 2.28: Percentage errors from fitting parameters simultaneously using the simplex algorithm on virtual patients P_1, ..., P_{50} generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, γ_V, β_{BI}, and β_{IB} are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

2.7 Identifiability analysis using the SAEM algorithm

In this section, we consider the fixed parameter set $\{\mu_{VT},\ \gamma_V,\ \beta_{BI},\ \beta_{IB}\}$ for which the results of the parameter fits from Section 2.6 had a mean error of 2.55%; the lowest out of all the fixed parameter sets that we fitted. In essence, the experiment is the same as with one key change made to the virtual patient datasets U_i. As it stands, a given U_i is exactly the matrix whose columns are the solution vectors for each compartment on the given timespan $\mathbf{t}$. In order to make use of the population parameter sampling from the conditional distribution in the SAEM algorithm, we duplicate each solution vector so that there are five copies and concatenate them together such that the matrix columns are now the same solution vectors but repeated five times. With this format, Monolix now considers each repeated set of vectors to constitute a different patient. Essentially, we change from considering the immune response within a single virtual patient to considering the immune response within a population of five identical virtual patients. While the new population size per fit of $N = 5$ is still not very large, we will see that this substantially improves the errors compared to what we previously observed under SAEM in Sections 2.4 and 2.5. Since the only change is adding copies of the numerical solutions to each matrix column, we do not need to regenerate the virtual patients from scratch. Instead, we simply re–define each solution matrix as

$$\mathcal{U}_i = \begin{pmatrix} U_i \\ U_i \\ U_i \\ U_i \\ U_i \end{pmatrix}$$

So each $\mathcal{U}_i$ is the solution matrix generated from the parameter vector $\mathbf{p}_i$ but with the solutions repeating five times. Given all of this, we can re–define the virtual patients $P_1,\ ...,\ P_{50}$ under

$$\mathcal{P}_i = (\mathbf{p}_i,\ \mathcal{U}_i)$$

Hence, the virtual patients that we now fit in Monolix are technically populations of identical patients which we consider as a single patient by taking the results of Monolix's population fit following SAEM. For the remainder of this study, with the exception of Section 3.5, we will consider the modified virtual patient population $\mathcal{P}_1,\ ...,\ \mathcal{P}_{50}$. We now repeat the experiment from Section 2.6 using this new population with the fixed parameter set $\{\mu_{VT},\ \gamma_V,\ \beta_{BI},\ \beta_{IB}\}$ and random effects re–enabled which switches the fitting algoritm back to SAEM. The experimental procedure is summarised as follows

1. Load the solution dataset $\mathcal{U}_i$ which is equivalent to U_i but with each solution repeated five times
2. Set all compartments as observations
3. Load the LNP model equations
4. Fix s_I, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Additionally, fix μ_{VT}, γ_V, β_{BI} and β_{IB} to their exact values from MATLAB with random effects disabled
6. Set all other parameters as variable and to be fitted under the maximum likelihood estimation with random effects enabled, which reverts the algorithm back to SAEM, and an initial value equal to the expected value of the respective uniform distribution
7. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space (5 iterations in most cases)
8. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
9. Run the SAEM algorithm which outputs the population parameter fits, that are now different from the individual patient fits as Monolix considers each repeated solution set as a distinct patient

The results are shown in Table 2.29. The mean error of the parameter fits under this revised method using SAEM is 1.98% which is an improvement over the mean error of 2.55% attained from the simplex algorithm seen in Table 2.24. We suspect that this is the result of taking better advantage of the nature of the SAEM algorithm with a larger populaiton size. We see that Monolix fits the overwhelming majority of parameters extremely close to their exact MATLAB values with only $8/50$ (16%) of patients showcasing errors $\geq 10\%$ and only $15/50$ (30%) of patients showcasing errors $\geq 5\%$. Crucially, we see that the errors that are noticeably high are clear outliers that are isolated to only a few patients. For example, out of the eight patients with errors exceeding 10%, six of them only have a single parameter that does so. This further supports the idea that Monolix propagates errors to certain parameters as a result of fitting others.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	4.03%	0.00%	1.43%	1.85%	1.58%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2	0.24%	0.00%	0.00%	0.00%	0.11%	0.00%	0.00%	0.00%	0.00%	11.54%	0.00%	0.00%
3	3.08%	0.00%	0.00%	1.96%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
4	7.62%	0.00%	7.14%	0.00%	2.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
5	0.00%	0.00%	1.00%	1.85%	1.82%	0.00%	0.00%	0.00%	0.00%	10.81%	0.00%	0.00%
6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
7	0.00%	0.00%	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
9	0.00%	0.00%	0.00%	0.00%	0.17%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
10	0.00%	0.00%	2.63%	2.02%	1.29%	0.00%	0.00%	0.00%	0.00%	5.13%	0.00%	0.00%
11	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12	0.38%	0.00%	0.00%	0.00%	0.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
13	5.71%	0.00%	69.23%	270.97%	2.90%	0.00%	8.33%	0.00%	0.00%	0.00%	0.00%	0.00%
14	0.00%	0.00%	0.00%	0.00%	1.34%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
15	1.99%	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
16	2.99%	0.00%	2.17%	0.00%	2.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
17	0.27%	0.00%	7.00%	0.00%	2.12%	0.00%	0.00%	1.69%	0.00%	1.96%	0.00%	0.00%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
19	3.69%	0.00%	1.37%	0.00%	0.18%	0.00%	0.00%	0.00%	0.00%	2.56%	0.00%	0.00%
20	6.62%	0.00%	2.38%	1.18%	1.53%	0.00%	0.00%	0.00%	0.00%	2.63%	0.00%	0.00%
21	0.00%	0.00%	1.39%	0.00%	0.19%	0.00%	0.00%	1.05%	0.00%	9.68%	0.00%	0.00%
22	0.00%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
23	0.57%	0.00%	0.00%	0.00%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
24	0.65%	0.00%	0.00%	1.06%	1.39%	0.00%	0.00%	0.00%	0.00%	2.22%	0.00%	0.00%
25	15.75%	0.00%	4.05%	0.00%	0.85%	0.00%	1.28%	0.00%	0.00%	2.94%	0.00%	0.00%
26	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
27	5.97%	0.00%	0.00%	0.00%	1.75%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
28	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
29	1.66%	0.00%	54.02%	16.67%	18.90%	0.00%	0.00%	3.57%	0.00%	0.00%	0.00%	0.00%
30	0.00%	0.00%	0.00%	0.00%	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
31	2.65%	0.00%	0.00%	1.03%	1.58%	0.00%	1.20%	0.00%	0.00%	6.45%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
33	2.22%	0.00%	1.15%	0.00%	0.38%	0.00%	0.00%	0.00%	0.00%	2.70%	0.00%	0.00%
34	0.00%	0.00%	4.11%	0.00%	1.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
35	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
37	0.40%	0.00%	0.00%	0.00%	0.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
38	0.00%	0.00%	2.56%	1.37%	0.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
39	0.00%	0.00%	0.00%	0.00%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	0.40%	0.00%	1.18%	0.00%	0.67%	0.00%	0.00%	1.79%	0.00%	2.22%	0.00%	0.00%
41	0.00%	0.00%	4.49%	0.00%	1.70%	0.00%	0.00%	0.00%	0.00%	45.71%	0.00%	0.00%
42	0.00%	0.00%	0.00%	0.00%	0.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
43	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
44	21.48%	0.00%	1.01%	0.00%	2.17%	0.00%	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%
45	1.54%	0.00%	0.00%	0.00%	1.32%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%	0.00%
46	0.00%	0.00%	4.00%	1.04%	0.08%	0.00%	0.00%	0.00%	0.00%	28.00%	0.00%	0.00%
47	3.28%	0.00%	0.00%	1.56%	2.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
48	0.00%	0.00%	2.04%	0.00%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
49	0.00%	0.00%	0.00%	0.00%	0.75%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
50	1.45%	0.00%	0.00%	1.22%	1.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Arithmetic Mean	1.90%	0.00%	3.56%	6.17%	1.21%	0.00%	0.25%	0.17%	0.00%	2.81%	0.00%	0.00%
Median	0.00%	0.00%	0.00%	0.00%	0.65%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 2.29: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on virtual patients $\mathcal{P}_1$, ..., $\mathcal{P}_{50}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, γ_V, β_{BI}, and β_{IB} are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

We briefly comment on the large outliers seen in the errors of the virtual patient $\mathcal{P}_{13}$ where the plasma B–cell priming rate μ_{TB} was fitted with an error of 69.23% and the antibody priming rate μ_{BA} was fitted with an error of 270.97%. As can be seen in Table 2.22, this patient, while having its parameters generated randomly like all the others, happened to have low priming rates and high decay rates for both vaccinated cells (V) and T-cells (T) as well as a high decay rate of plasma B–cells (B). The full parameter vector $\mathbf{p}_{13}$ can be seen in Table 2.30.

Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
$\mathbf{p}_{13}$	0.7	0.8	0.13	0.62	5.86	0.11	0.12	0.5	0.041	0.04	2.96	0.0016	1000	157.19	135.62

Table 2.30: Parameter vector $\mathbf{p}_{13}$ of virtual patient $\mathcal{P}_{13}$ demonstrating noticeably low priming rates and high decay rates for vaccinated cells (V) and T-cells (T) as well as a high decay rate of plasma B–cells (B).

Comparing the parameter values to the distributions in Table 2.5, we can highlight the following. The priming rates $\mu_{LV} = 0.7$ and $\mu_{VT} = 0.8$ are low given that the minimum and maximum values of their uniform distributions are

$$\begin{aligned}(p_{\min},\ p_{\max})(\mu_{LV}) &= (0.153,\ 4.268) \\ (p_{\min},\ p_{\max})(\mu_{VT}) &= (0.675,\ 29.524)\end{aligned} \tag{2.31a}$$

Similarly the decay rates $\gamma_V = 0.11$, $\gamma_T = 0.12$ and $\gamma_B = 0.5$ are high given that

$$\begin{aligned}(p_{\min},\ p_{\max})(\gamma_V) &= (0.0513,\ 0.143) \\ (p_{\min},\ p_{\max})(\gamma_T) &= (0.0396,\ 0.121) \\ (p_{\min},\ p_{\max})(\gamma_B) &= (0.0234,\ 0.594)\end{aligned} \tag{2.31b}$$

The result is that the upstream compartments from the antibodies (A) have much lower peaks than is typical and this results in the rare instance where the virtual patient does not exhibit an antibody peak. Instead, the antibody concentration never exceeds the initial value $A_0 = 157.19$ and decays for the entire timespan with the exception of a slight jump around $t = 30$ days. The full plot of the antibody solution curve can be seen in Figure 2.32. We suspect then that the large errors in the parameter fits for patient $\mathcal{P}_{13}$ are precisely due to the fact that there is no visible peak in the antibodies. This may reflect the dynamics of a weak immune system however it is difficult to say this conclusively. Separate studies into the identifiability of model parameters in weak systems may yield greater insights.

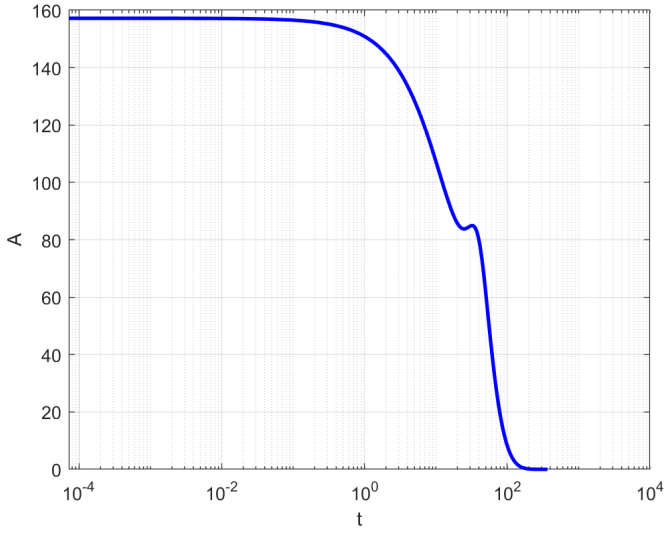


Figure 2.32: MATLAB plot of the antibodies (A) for patient $\mathcal{P}_{13}$ demonstrating a lack of an global antibody peak following vaccine administration

Moving back to the full sets of parameter fits, the results of this section as well as those of Section 2.6 have demonstrated two main points. Firstly, we have shown that when no parameters are fixed, the results of Monolix do not align with the structural identifiability implied in DAISY as we would expect minimal errors in those cases. Secondly, we have shown that fixing certain parameters can greatly impact the accuracy of the parameter fits and, crucially, that not all other parameters need to be fixed in order to accurately fit a given parameter as was implied in Tables 2.12 and 2.18. Finally, we have shown that by modifying the structure of individual patient datasets such that they can be considered as populations, the SAEM algorithm offers improved fits compared to the simplex algorithm. Expanding the population of identical patients to be larger than $N = 5$ may further improve the results albeit at the cost of computational complexity. We suggest this as a possible future study. To conclude, we propose, as before, that the need for fixing certain parameters in order to obtain accurate fits stems from practical identifiability issues involving the distribution of data points and the parameter values themselves. Therefore, we now perform a more detailed analysis into the analytic solutions of the LNP model.

2.8 Identifiability analysis from the analytic solutions

In this section, we take a more detailed look at the analytic solutions of the LNP model with the goal of identifying why some parameters are fitted poorly and how fixing some parameters can make it easier to identify others. The solutions for L, V, T, and A are the same as equations 2.3a

– 2.3d with the γ_L term removed. As mentioned in Section 2.1, the full solutions are shown in Appendix A. For completeness, we list them here in their final closed forms

$$L(t) = e^{-\mu_{LV} t} \tag{2.33a}$$

$$V(t) = \frac{\mu_{LV}}{\gamma_V - \mu_{LV}}(e^{-\mu_{LV} t} - e^{-\gamma_V t}) \tag{2.33b}$$

$$T(t) = \frac{\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}\left(\frac{e^{-\mu_{LV} t} - e^{-\gamma_T t}}{\gamma_T - \mu_{LV}} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V}\right) \tag{2.33c}$$

$$A(t) = e^{-\gamma_A t}\left[A_0 + \mu_{BA}\int_0^t e^{\gamma_A s} B(s)\, ds\right] \tag{2.33d}$$

Furthermore, while B and I are represented by the non–linear equations 2.20d and 2.20f, we can attempt to infer the impact of the parameter values on the linear growth and decay terms by setting $\beta_{BI} = \beta_{IB} = 0$. Similar to the use of $\widetilde{C}$ and $\widetilde{F}$ in equations 2.17f and 2.17g, we write the linearised equations with $\widetilde{B}$ and $\widetilde{I}$

$$\frac{d\widetilde{B}}{dt} = \mu_{TB} T - \gamma_B \widetilde{B} \tag{2.34a}$$

$$\frac{d\widetilde{I}}{dt} = \mu_{TI} T - \gamma_I \widetilde{I} \tag{2.34b}$$

The initial conditions are once again the same with $\widetilde{B}(0) = 0$ and $\widetilde{I}(0) = I_0$. These equations now continue the linear cascade from T and can be solved in the same way. The full solutions are once again shown in Appendix Awith the closed forms as follows

$$\widetilde{B}(t) = \frac{\mu_{TB}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}\left[\frac{e^{-\mu_{LV} t} - e^{-\gamma_B t}}{(\gamma_T - \mu_{LV})(\gamma_B - \mu_{LV})} + \frac{e^{-\gamma_T t} - e^{-\gamma_B t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_T)}\right. \tag{2.35a}$$

$$\left. - \frac{e^{-\gamma_T t} - e^{-\gamma_B t}}{(\gamma_T - \mu_{LV})(\gamma_B - \gamma_T)} - \frac{e^{-\gamma_V t} - e^{-\gamma_B t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_V)}\right]$$

$$\widetilde{I}(t) = \frac{\mu_{TI}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left[\frac{e^{-\mu_{LV}t} - e^{-\gamma_I t}}{(\gamma_T - \mu_{LV})(\gamma_I - \mu_{LV})} + \frac{e^{-\gamma_T t} - e^{-\gamma_I t}}{(\gamma_T - \gamma_V)(\gamma_I - \gamma_T)} \right. \tag{2.35b}$$

$$\left. - \frac{e^{-\gamma_T t} - e^{-\gamma_I t}}{(\gamma_T - \mu_{LV})(\gamma_I - \gamma_T)} - \frac{e^{-\gamma_V t} - e^{-\gamma_I t}}{(\gamma_T - \gamma_V)(\gamma_I - \gamma_V)} \right] + I_0 e^{-\gamma_I t}$$

Given $\widetilde{B}(t)$, we can do the same for A and write a closed form for 2.33d with $A(t)$ and $B(s)$ replaced by $\widetilde{A}(t)$ and $\widetilde{B}(s)$ respectively

$$\widetilde{A}(t) = \frac{\mu_{BA}\mu_{TB}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left[\frac{e^{-\mu_{LV}t} - e^{-\gamma_A t}}{(\gamma_T - \mu_{LV})(\gamma_B - \mu_{LV})(\gamma_A - \mu_{LV})} + \frac{e^{-\gamma_T t} - e^{-\gamma_A t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_T)(\gamma_A - \gamma_T)} \right.$$

$$+ \frac{e^{-\gamma_B t} - e^{-\gamma_A t}}{(\gamma_T - \mu_{LV})(\gamma_B - \gamma_T)(\gamma_A - \gamma_B)} + \frac{e^{-\gamma_B t} - e^{-\gamma_A t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_V)(\gamma_A - \gamma_B)}$$

$$- \frac{e^{-\gamma_B t} - e^{-\gamma_A t}}{(\gamma_T - \mu_{LV})(\gamma_B - \mu_{LV})(\gamma_A - \gamma_B)} - \frac{e^{-\gamma_B t} - e^{-\gamma_A t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_T)(\gamma_A - \gamma_B)}$$

$$\left. - \frac{e^{-\gamma_T t} - e^{-\gamma_A t}}{(\gamma_T - \mu_{LV})(\gamma_B - \gamma_T)(\gamma_A - \gamma_T)} - \frac{e^{-\gamma_V t} - e^{-\gamma_A t}}{(\gamma_T - \gamma_V)(\gamma_B - \gamma_V)(\gamma_A - \gamma_V)} \right]$$

$$+ A_0 e^{-\gamma_A t} \tag{2.35c}$$

We briefly highlight the presence of potential singularities in equations 2.33c and 2.35a – 2.35c. Namely we are concerned with the terms $\gamma_i - \gamma_j$ which divide the exponential differences. Referring back to Table 2.5, we see that every decay rate γ_i has its distribution overlap with at least one other decay rate γ_j. Furthermore, since all the decay rates are between 0 and 1, they are generated with 2 significant figures of accuracy. Therefore, it is possible for a virtual patient to be generated with $\gamma_i = \gamma_j$ which results in a division by zero. This reflects the repeated eigenvalue case for linear equations since, within the LNP model equations, all of the decay rates multiply their respective compartment variables. In this case, we can rectify the presence of the singularity by simply re–solving the analytic solutions with the two decay rates merged into one, $\gamma_i = \gamma_j = \gamma$. While technically a singularity like this can exist in equations 2.35a – 2.35c, they all stem from modifying

the model to make B and I linear. These solutions will not be considered beyond this section and so we only concern ourselves with the case of $\gamma_V = \gamma_T = \gamma$ in equation 2.33c since it stems from the LNP model itself. Given this, we re–solve the analytic solutions for V and T. In the case of V, since only γ_V is present, the solution is the same as 2.33b with the decay parameter changed to γ

$$V(t) = \frac{\mu_{LV}}{\gamma - \mu_{LV}}(e^{-\mu_{LV}t} - e^{-\gamma t}) \tag{2.36a}$$

For T, the equality results in a slightly different solution. The full analysis is once again shown in Appendix Awith the final closed form shown below

$$T(t) = \frac{\mu_{VT}\mu_{LV}}{(\gamma - \mu_{LV})^2}\left(e^{-\mu_{LV}t} + e^{-\gamma t} - (\gamma - \mu_{LV})te^{-\gamma t}\right) \tag{2.36b}$$

Having addressed the issue of singularities, we now consider the effects of the fixed parameter tested thus far given the analytic solutions. We acknowledge that the fixing of μ_{LV} in Tables 2.26 and 2.27 was likely not necessary and instead could have been replaced with the fixing of μ_{VT} as can clearly be seen by the improvement in errors from Table 2.27 to Table 2.28. Since μ_{LV} is now the unique decay rate of L, we expect to be able to capture its value, even when given limited decay curve data, and indeed it had the second lowest mean error and lowest median error out of all fitted parameters in the control case. We suspect that the higher errors of μ_{LV} visible on a handful of patients are due to the propagation of errors from parameters being fitted downstream. We now consider μ_{VT} specifically as it has the second highest mean error and is a parameter for which several patients exhibited divergent fits with errors exceeding 1000% while also being a crucial parameter that drives the immune response seeing as the cascading linear dynamics change to being non–linear as a result of T priming B and I. We see that in both the analytic solution for T in equation 2.33c and the linearised solutions for B, A, and I in equations 2.35a – 2.35c, μ_{VT} only ever appears in the scalar coefficient ratios

$$\frac{\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}, \quad \frac{\mu_{TB}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}, \quad \frac{\mu_{TI}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}, \quad \frac{\mu_{BA}\mu_{TB}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}. \tag{2.37}$$

Since in this chapter we have considered all compartments as observed state variables, all of the downstream effects from the more complicated ratios in equation 2.37 are present. Therefore, we consider μ_{VT} as a reasonable parameter to fix given that it is upstream from said effects which make it harder to isolate. This will be discussed further in Chapter 3 when only the downstream compartments will be observed. Referring back to Table 2.23, we can see that in the cases where

μ_{VT} has a high error ($\geq 100\%$), at least one of the other priming rates μ_{ij} or γ_V also have high errors. The only exception is patient P_{35} which does however have a high error on γ_T; further suggesting that the fitting of μ_{VT} is more dependent on the fitting of other parameters compared to those present in compartments further downstream. We hence justify the fixing of μ_{VT} as not only does it reduce the variability of the ratio coefficients but also, due to it being upstream of the non–linear dynamics, we suspect that fixing it has a greater impact on the accuracy of the downstream fits compared to the upstream impact of fixing μ_{TB} or μ_{TI}. This justification is supported by the results of Tables 2.25, 2.28, and 2.29.

Our justification for the fixing of γ_V comes from the fact that it had some of the highest mean and median errors from the decay rates γ_i in the control case. Furthermore, similar to μ_{VT}, it is upstream from the non–linear dynamics. Therefore, within equations 2.33b and 2.33c as well as 2.35a – 2.35c, γ_V appears in exponential difference terms of the form $\pm(e^{-kt} - e^{-\gamma_V t})$. Depending on the individual parameter values taken on by a patient, these exponential terms may become subdominant compared to the other term in the difference and hence harder to identify. The fixing of the non–linear autocatalytic rates β_{BI} and β_{IB} cannot be justified given the linear solutions since they are not present. The non–linearity itself adds complexity which can make them harder to identify. Additionally, in the case of β_{IB}, there may be a similar issue to that of γ_L which we discussed in Sections 2.3 and 2.5 where the small size of the parameter, which the algorithms in Monolix seemingly struggle with, may be the result of it not being identifiable given the laboratory data when it was fitted in Korosec et al. [27].

Going back to the issue with not having sufficient decay data on certain patients to capture the effect of γ_V. The linear cascade of model solutions 2.33a, 2.33b, and 2.33c allows for us to take advantage of the model structure. Since those compartments de–couple, we can fit them individually and then fit the non–linear compartments together. By doing this, the dimension of the parameter spaces when fitting the upstream compartments is reduced substantially which greatly simplifies the calculations performed by the SAEM algorithm. The hope is that by employing the model structure in such a way, we can improve the accuracy of the fits without necessarily needing to fix as many parameters. We now consider this in the next section.

2.9 Identifiability analysis by model compartment

We finish the analysis of this chapter by now considering the case where the model compartments are fitted successively based on the flow of the dynamics activating each compartment starting from L and moving downstream towards A. For this, we perform a modified version of the experiment

in Section 2.7 by taking ten virtual patients that exhibited some of the highest errors in the simplex algorithm control case seen in Table 2.23 but with the improved fitting method for SAEM where each dataset is repeated five times. The selected patients from $\mathcal{P}_1$, ..., $\mathcal{P}_{50}$ are the following

$$\{\mathcal{P}_{14},\ \mathcal{P}_{16},\ \mathcal{P}_{18},\ \mathcal{P}_{20},\ \mathcal{P}_{30},\ \mathcal{P}_{32},\ \mathcal{P}_{34},\ \mathcal{P}_{35},\ \mathcal{P}_{36},\ \mathcal{P}_{49}\}$$

For these compartmental fits, we start with fitting just L and compare the fitted parameter values to the exact MATLAB values like before. Then we fit V by providing its analytic solution instead of its differential equation and fix the L parameters to their exact values by assuming that we have already identified the parameters of the upstream compartments correctly. We will justify setting the upstream parameters to their exact values later in this section given the results of the fits. We repeat this process for T. We provide its analytic solution and fix both the L and V parameters to their exact values under the same assumption. Then we reach the non–linear compartments B and I which are coupled. So we fit these together by providing their differential equations with the analytic solution of T substituted in and the L, V, and T parameters fixed to their exact values. Finally, we fit A given its differential equation and the modified B and I equations containing the solution of T with the L, V, T, B, and I parameters fixed to their exact values. Starting with L, which in this case is just fitting a single exponential decay parameter μ_{LV}, for each patient

1. Load the solution dataset $\mathcal{U}_i$
2. Set L as the observed data
3. Load equation 2.20a as the model equation for L
4. Set the sole parameter, μ_{LV}, as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution
5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space
6. Set the posterior distribution for the data on L to a normal distribution (Monolix default)
7. Run the SAEM algorithm which outputs a population parameter fit that functions as an individual fit

For the remaining compartments, we take advantage of the analytic solutions by substituting them into the linear cascade which reduces the number of equations in the model provided to Monolix.

With less compartments to be considered, the hope is that the parameter fitting procedure is greatly simplified. So for V, we substitute 2.33a for L which means that we input

$$\frac{dV}{dt} = \mu_{LV} e^{-\mu_{LV} t} - \gamma_V V \tag{2.38a}$$

We recall that the presence of μ_{LV} as the priming rate of V is the unique case of a parameter appearing in more than one compartment. Since it was already fitted when considering just L, we assume its value is known exactly and fix it to its exact MATLAB value. Hence, in this case, we are once again fitting a single parameter, γ_V. We will provide a full justification for the fixing later in this section when we discuss the results of the fits. Now given this, we fit V for each patient as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set V as the observed data
3. Load equation 2.38a as the model equation for V
4. Fix μ_{LV} to its exact MATLAB value, and set γ_V as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution
5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space
6. Set the posterior distribution for the data on V to a normal distribution (Monolix default)
7. Run the SAEM algorithm which outputs a population parameter fit that functions as an individual fit

The procedure is similar for T in that we substitute 2.33b for V meaning that we input

$$\frac{dT}{dt} = \frac{\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}(e^{-\mu_{LV} t} - e^{-\gamma_V t}) - \gamma_T T \tag{2.38b}$$

Once again, we assume that the upstream compartments L and V have already been fitted and their parameters are known exactly allowing us to fix μ_{LV} and γ_V to their MATLAB values. Then we can fit T as follows

1. Load the solution dataset $\mathcal{U}_i$

2. Set T as the observed data

3. Load equation 2.38b as the model equation for T

4. Fix μ_{LV}, γ_V to their exact MATLAB values, and set μ_{VT}, γ_T as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution

5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space

6. Set the posterior distribution for the data on T to a normal distribution (Monolix default)

7. Run the SAEM algorithm which outputs the population parameter fits that function as individual fits

The linear dynamics stop with T priming B and I which are coupled being regulated by the non–linear autocatalytic terms driven by β_{BI} and β_{IB}. So we now fit B and I together with 2.33c substituted for for T. This gives the equations

$$\frac{dB}{dt} = \frac{\mu_{TB}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left(\frac{e^{-\mu_{LV}t} - e^{-\gamma_T t}}{\gamma_T - \mu_{LV}} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V} \right) + \beta_{BI} \left(\frac{I}{I + s_I} \right) B - \gamma_B B \tag{2.38c}$$

$$\frac{dI}{dt} = \frac{\mu_{TI}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left(\frac{e^{-\mu_{LV}t} - e^{-\gamma_T t}}{\gamma_T - \mu_{LV}} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V} \right) - \beta_{IB} IB - \gamma_I I \tag{2.38d}$$

We recall from Section 2.8 that the $\gamma_V - \gamma_T$ terms can result in a singularity due to the fact that the distributions overlap making it possible for a virtual patient to have $\gamma_V = \gamma_T$ which results in a division by zero in equation 2.33c. In fact, this is exactly what happens with $\mathcal{P}_{49}$ where $\gamma_V = \gamma_T = 0.12$. This is the case for which we derived the modified analytic solution for T given by equation 2.36b and so for $\mathcal{P}_{49}$ we substitute that into the equations for B and I instead which gives

$$\frac{dB}{dt} = \frac{\mu_{TB}\mu_{VT}\mu_{LV}}{(\gamma - \mu_{LV})^2} \left(e^{-\mu_{LV}t} + e^{-\gamma t} - (\gamma - \mu_{LV}) t e^{-\gamma t} \right) + \beta_{BI} \left(\frac{I}{I + s_I} \right) B - \gamma_B B \tag{2.39a}$$

$$\frac{dI}{dt} = \frac{\mu_{TI}\mu_{VT}\mu_{LV}}{(\gamma - \mu_{LV})^2} \left(e^{-\mu_{LV}t} + e^{-\gamma t} - (\gamma - \mu_{LV}) t e^{-\gamma t} \right) - \beta_{IB} IB - \gamma_I I \tag{2.39b}$$

Again we assume that the upstream compartments L, V, and T have already been fitted with their parameter values known exactly. Hence, we fix μ_{LV}, μ_{VT}, γ_V, and γ_T to their MATLAB values. Also note that s_I, and I_0 are already fixed from before. Therefore, B and I are fitted together as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set B and I as the observed data
3. Load equation 2.38c as the model equation for B and equation 2.38d as the model equation for I
    - If $\gamma_V = \gamma_T$ then load equations 2.39a and 2.39b for B and I instead
4. Fix μ_{LV}, μ_{VT}, γ_V, γ_T to their exact MATLAB values (s_I, I_0 are already fixed from before), and set μ_{TB}, μ_{TI}, γ_B, γ_I, β_{BI}, β_{IB} as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution
5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space
6. Set the posterior distributions for data on B and I to normal distributions (Monolix default)
7. Run the SAEM algorithm which outputs the population parameter fits that function as individual fits

Finally, we consider A which has a linear equation however it is primed by B which is non–linear and so we cannot substitute an analytic solution. Furthermore, B is coupled with I through the regulatory autocatalytic process. So instead, we set A as the observed data provided through equation 2.20e and also provide equations 2.38c and 2.38d for B and I respectively due to the non–linearity and coupling. Once again, in the case of $\mathcal{P}_{49}$ where $\gamma_V = \gamma_T$, we instead provide equations 2.39a and 2.39b. Just like before, we now assume that all the upstream compartments L, V, T, B, and I have already been fitted with their parameters known exactly. So we fix μ_{LV}, μ_{VT}, μ_{TB}, μ_{TI}, γ_V, γ_T, γ_B, γ_I, β_{BI}, and β_{IB}. Again, s_I and I_0 are fixed from before with A_0 now also present and fixed as it was. Thus, A is fitted as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set A as the observed data

3. Load equation 2.38c as the model equation for B, equation 2.38d as the model equation for I, and equation 2.20e as the model equation for A

    - If $\gamma_V = \gamma_T$ then load equations 2.39a and 2.39b for B and I instead

4. Fix μ_{LV}, μ_{VT}, μ_{TB}, μ_{TI}, γ_V, γ_T, γ_B, γ_I, β_{BI}, β_{IB} to their exact MATLAB values (s_I, A_0, I_0 are already fixed from before), and set μ_{BA}, γ_A as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution

5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space

6. Set the posterior distribution for the data on A to a normal distribution (Monolix default)

7. Run the SAEM algorithm which outputs the population parameter fits that function as individual fits

To quickly recap, we fit the compartments of the LNP model successively in Monolix given the virtual patient set $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$. We first fit L, then V, then T, then B and I together, and finally A. In each case, we substitute in the analytic solutions of L, V, and T wherever possible to reduce the number of compartments being fitted with parameters from the previously fitted upstream compartments being fixed to their exact MATLAB values. For $\mathcal{P}_{49}$, we use a modified solution of T when fitting B, I, A to account for the fact that $\gamma_V = \gamma_T$. The results of these compartmental fits are shown in Table 2.40. We see a substantial improvement compared to the control case in Table 2.23 and in fact most parameters have errors that are just as low if not lower than the best case with fixed parameters and SAEM seen in Table 2.29. The key difference here is that we have fixed no parameters aside from s_I, A_0, and I_0 which were already fixed in the previous cases. Furthermore, unlike the previous cases where the distribution of higher errors across parameters seemed inconsistent, in this case, the higher errors are consistently distributed on the same parameters; these being μ_{TB} and β_{IB}. Therefore, we have demonstrated that by taking advantage of the model structure allowing for us to make use of the analytic solutions, the errors of the parameter fits are substantially improved when fitting the compartments successively rather than all at once.

Observing the error values in more detail, we note that, with the exception of μ_{TB} and β_{IB} as mentioned already, most parameter values are fitted exactly while those that are not still have errors less than 10%. Additionally, the mean errors do not exceed 2%. This helps to justify our choice at the beginning of this section where previously fitted upstream parameters are fixed to

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
14	0.00%	0.07%	32.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	46.15%
16	2.56%	9.34%	23.91%	0.00%	0.05%	0.00%	1.00%	0.00%	0.00%	0.00%	0.00%	32.26%
18	0.00%	0.03%	16.67%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	18.18%
20	8.82%	3.38%	17.86%	0.00%	0.04%	0.00%	2.60%	0.00%	0.00%	0.00%	0.00%	22.58%
30	0.00%	0.04%	16.49%	2.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	19.44%
32	2.51%	0.00%	14.29%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.05%
34	0.00%	0.00%	19.18%	0.00%	0.05%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.81%
35	3.17%	0.00%	15.49%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.05%
36	0.00%	0.12%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%	21.21%
49	0.00%	6.89%	21.43%	0.00%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	30.56%
Arithmetic Mean	1.71%	1.99%	19.43%	0.24%	0.03%	0.00%	0.36%	0.00%	0.00%	0.00%	0.07%	25.83%
Median	0.00%	0.05%	17.26%	0.00%	0.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.90%

Table 2.40: Percentage errors from fitting parameters on the virtual patient set $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. The model compartments are fitted separately wherever de–coupling is possible and the analytic solutions are provided whenever they are available. First L is fitted as the only observed compartment, followed by V as the only observed compartment, then T as the only observed compartment, then B and I together as the only observed compartments, and finally A being an observed compartment with B and I also observed due to the non–linearity. In each case, any upstream parameters from previously fitted compartments are fixed to their exact values. Also shown are the mean and median errors for each parameter. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

their exact values rather than their fitted values. Since the fitted values are very close to the exact MATLAB values, we assume that any impact on the accuracy of the fits from not fixing upstream parameters to the fitted values is minimal. The only compartment for which we may expect a more noticeable impact is A, since both μ_{TB} and β_{IB} are upstream from it and demonstrated higher errors. Now we can consider fixing one of μ_{TB} or β_{IB} with the hope that the errors of the other parameter are reduced. Out of these two, we choose to fix β_{IB} due to it being non–linear, having a higher mean error than μ_{TB} in Table 2.40, and having a lower variance given the distributions in Table 2.5. Notably, here we are fixing just one parameter, not including s_I, A_0, and I_0 which are always fixed, compared to the four parameters fixed earlier in Table 2.29. The structure of the experiment is the same and so the only change is in the initial setup when we fit B and I. Therefore, we only need to fit those two compartments again as the setups for L, V, T, and A are the same. Fixing β_{IB}, the experimental procedure is as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set B and I as the observed data
3. Load equation 2.38c as the model equation for B and equation 2.38d as the model equation for I

- If $\gamma_V = \gamma_T$ then load equations 2.39a and 2.39b for B and I instead

4. Fix μ_{LV}, μ_{VT}, γ_V, γ_T, β_{IB} to their exact MATLAB values (s_I, I_0 are already fixed from before), and set μ_{TB}, μ_{TI}, γ_B, γ_I, β_{BI} as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution

5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space

6. Set the posterior distributions for data on B and I to normal distributions (Monolix default)

7. Run the SAEM algorithm which outputs the population parameter fits that function as individual fits

The results are shown in Table 2.41. We see that the fixing of β_{IB} results in a substantial improvement in the errors of μ_{TB} with mean and median errors of 4.01% and 3.63% respectively compared to 19.43% and 17.26% in Table 2.40. Crucially, the errors in the other parameters do not seem to be affected by propagation resulting from the fixing. This is a key difference from the behaviour we observed in the results of Section 2.6 and further supports the idea that it is better to take advantage of the analytic solutions of observed compartments allowing for us to fit them successively.

Patient ID \ Parameter	μ_{TB}	μ_{TI}	γ_B	γ_I	β_{BI}	β_{IB}
14	7.69%	0.35%	1.82%	0.00%	2.05%	0.00%
16	6.52%	0.21%	0.00%	0.00%	1.23%	0.00%
18	0.00%	0.13%	0.00%	0.00%	0.93%	0.00%
20	1.19%	0.04%	0.00%	0.00%	0.43%	0.00%
30	2.06%	0.63%	0.00%	0.00%	0.74%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.31%	0.00%
34	8.22%	0.27%	2.33%	0.00%	1.68%	0.00%
35	4.23%	0.15%	0.00%	0.00%	0.85%	0.00%
36	3.03%	0.12%	0.00%	0.00%	0.64%	0.00%
49	7.14%	0.27%	0.00%	0.00%	0.75%	0.00%
Arithmetic Mean	4.01%	0.22%	0.41%	0.00%	0.96%	0.00%
Median	3.63%	0.18%	0.00%	0.00%	0.80%	0.00%

Table 2.41: Percentage errors from fitting parameters of B and I on the virtual patient set $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, β_{IB} is fixed to its exact value. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Finally, we can consider one more modification to the experiment. When fitting A earlier, we still had to provide the model equations for B and I since there is no analytic solution for B to substitute into the $\mu_{BA}B$ term in equation 2.20e unlike the case for V, T, B, and I. To avoid needing to consider B and I data twice, we can simply fit all three compartments together instead of fitting A separately. We also maintain the fixing of β_{IB}. Therefore, we have the following experimental procedure

1. Load the solution dataset $\mathcal{U}_i$
2. Set B, A, and I as the observed data
3. Load equation 2.38c as the model equation for B, equation 2.20e as the equation for A, and equation 2.38d as the model equation for I
    - If $\gamma_V = \gamma_T$ then load equations 2.39a and 2.39b for B and I instead
4. Fix μ_{LV}, μ_{VT}, γ_V, γ_T, β_{IB} to their exact MATLAB values (s_I, A_0, I_0 are already fixed from before), and set μ_{TB}, μ_{BA}, μ_{TI}, γ_B, γ_A, γ_I, β_{BI} as variable and to be fitted with SAEM under the maximum likelihood estimation with an initial value equal to the expected value of the respective uniform distribution
5. Run the auto–init procedure iteratively until the variable initial parameter value has stabilised to a certain neighbourhood in the parameter space
6. Set the posterior distributions for data on B, A, and I to normal distributions (Monolix default)
7. Run the SAEM algorithm which outputs the population parameter fits that function as individual fits

The results are shown in Table 2.42. We observe further improvements in the errors of μ_{TB} with the mean reduced to 2.46% and the median reduced to 0.60%. Furthermore there is no indication that errors are propagating to A as a result of the fixing since the errors of μ_{BA} and γ_A are the same as those in Table 2.40. Hence, we have demonstrated that when fitting compartments successively, we can reduce the number of fits required by pairing downstream compartments with the non–linear compartments that they are activated by.

Before concluding this chapter, we highlight an interesting artifact observed in the parameter fits that was also observed in the testing of Sections 2.6 and 2.7 but is more clearly noticeable here.

Patient ID \ Parameter	μ_{TB}	μ_{BA}	μ_{TI}	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
14	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.41%	0.00%
16	4.35%	0.00%	0.95%	0.00%	0.00%	0.00%	0.93%	0.00%
18	0.00%	0.00%	0.13%	0.00%	0.00%	0.00%	0.31%	0.00%
20	1.19%	0.00%	0.21%	0.00%	0.00%	0.00%	0.43%	0.00%
30	2.06%	2.41%	1.90%	0.00%	0.00%	0.00%	0.49%	0.00%
32	0.00%	0.00%	0.04%	0.00%	0.00%	0.00%	0.31%	0.00%
34	2.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%	0.00%
35	0.00%	0.00%	0.05%	0.00%	0.00%	0.00%	0.21%	0.00%
36	0.00%	0.00%	0.12%	0.00%	0.00%	0.00%	0.21%	0.00%
49	14.29%	0.00%	1.51%	4.35%	0.00%	0.00%	3.74%	0.00%
Arithmetic Mean	2.46%	0.24%	0.50%	0.43%	0.00%	0.00%	0.75%	0.00%
Median	0.60%	0.00%	0.13%	0.00%	0.00%	0.00%	0.42%	0.00%

Table 2.42: Percentage errors from fitting parameters of B, A, and I together on the virtual patient set $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, β_{IB} is fixed to its exact value. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

So far we have not discussed in detail the results of the iterative auto–init procedure from which we attained the initial estimates of the parameters prior to fitting them. We now provide some further insights. As mentioned previously in the experimental summaries, for all virtual patients, aside from a few exceptions, we ran five iterations of the auto–init algorithm which was enough to have the initial parameter values converge to what Monolix believes to be a neighbourhood of the maximum likelihood estimation in the parameter space. In these cases, applying more iterations would no longer display noticeable changes in the initial values. When we began testing on the virtual patient population $\{P_1, ..., P_{50}\}$ with the simplex algorithm and fixed parameter sets, we noticed that more often than not, the auto–init procedure would have the initial parameter estimates converge to be very close to or equal to the exact MATLAB values. However, after running the parameter fits with the simplex algorithm, the values would diverge away from the initial estimates resulting in the final parameter fits having a higher error than the auto–init initial estimates. This same behaviour was also observed when we moved to testing on the modified virtual patient population $\{\mathcal{P}_1, ..., \mathcal{P}_{50}\}$ with the SAEM algorithm. An example of this is best seen when we look at the errors of the auto–init estimates for μ_{TB} and β_{IB} in Table 2.43 prior to attaining the SAEM results of Table 2.40.

Patient ID \ Parameter	μ_{TB}	β_{IB}
14	0.77%	2.31%
16	0.00%	0.65%
18	1.67%	15.45%
20	1.67%	6.77%
30	0.21%	5.28%
32	0.71%	4.74%
34	0.68%	6.13%
35	3.24%	3.68%
36	0.91%	4.85%
49	0.00%	1.11%
Arithmetic Mean	0.99%	5.10%
Median	0.74%	4.79%

Table 2.43: Percentage errors for μ_{TB} and β_{IB} of the initial parameter estimates attained through the auto–init algorithm prior to running SAEM which gave the results of Table 2.40.

As we can clearly see, all of the parameter estimates of μ_{TB} and β_{IB} have a lower error following auto–init compared to the results of the fits using SAEM. For this we refer back to the outline of the auto–init algorithm in Section 1.3. We recall that, after defining starting initial values for each parameter which in this case are the expected values (midpoints) of the parameter distributions, when the run button is clicked, two optimisation steps are initialised. First, in the global step, a weight sum least squares method outputs optimal parameter values given the provided datasets. Then in the local step, these values are provided as the starting points for the Nelder–Mead algorithm which produces a further optimised set of values. These become the new initial values for the next run of auto–init. It is difficult to conclusively say why this results in more accurate parameter fits compared to running the SAEM algorithm following auto–init. However we speculate that since the SAEM algorithm utilises posterior distributions which add an extra layer of randomness, under certain conditions, the optimal values attained from auto–init may not be re–identified if the iterates move further away in the parameter space.

We come to the following conclusions for this chapter. Firstly, it is clear that while the model may be structurally identifiable, the particular distributions of its parameters and the methods used to estimate them result in instances where, even given large datasets and observing all compartments, the individual parameter values cannot be correctly identified. We can speculate that such issues stem from both a lack of practical identifiability as well as some fitting algorithms just not being well suited to particular datasets. Secondly, we have demonstrated that by fixing model parameters, the estimations from the parameter fits can be more accurate however errors may propagate across compartments which may require fixing additional parameters. Finally, we have demonstrated that if the model structure allows for the analytic solutions of certain compartments to be attained, then

the number of fixed parameters required to attain much lower errors that are more consistently distributed is reduced. Of course, in the case of the LNP model, we could only utilise the analytic solutions as inputs when setting up the experiments in Monolix because all compartments were observed. However, this is almost never the case when it comes to clinical trials. For example, Korosec et al. [27], from whom we derived the parameter distributions, estimated the parameter values given individual data from only antibodies (A) and interleukin (I). In such cases, we cannot resort to inputting analytic solutions because we no longer observe the linear compartments L, V, and T. Hence, our considerations of fixed parameter sets in Sections 2.6 and 2.7 play an important role in informing us on the effect of fixed parameters on the errors when all compartments have to be fitted together. This now becomes relevant in Chapter 3 where we consider exactly those said cases where data from the linear compartments is no longer observed.

Chapter 3

In this chapter, we perform a similar parameter identifiability analysis to Chapter 2 using the LNP model defined in Section 2.6 and the modified virtual patient set $\mathcal{P}_1$, ..., $\mathcal{P}_{50}$ defined in Section 2.7. This time we consider cases where not all compartments of the model are observed as state variables. We begin by considering the case where the observed compartments are the plasma B–cells (B), antibodies (A), and interleukin (I). This initial choice stems from A and I being observed in Korosec et al. [27] and B being the remaining non–linear compartment that is coupled with I. We repeat the differential algebra analysis of DAISY outlined in Section 2.2 with this reduced compartment set. The model is no longer globally identifiable and so this time based on inferences from the DAISY results, we estimate the parameters in Monolix under different fixed parameter sets (Section 3.1). We do the same while observing only A and I as state variables which is equivalent to the case of Korosec et al. We also briefly comment on the case where only A is observed as a state variable since a lack of structural identifiability data from DAISY gives less concrete inferences on what parameters to fix (Section 3.2). In both of these sections, certain parameters become more identifiable when others are fixed. We discuss in more detail why this occurs given the effects of fixed parameters on the differential algebra outputs from DAISY (Section 3.3). The interleukin saturation s_I is indicated by DAISY to be structurally identifiable. Since this parameter was fixed from the early experiments of Chapter 2 as well as in Korosec et al., we consider the case where it is unfixed and the effect that it has on the parameter estimations in Monolix (Section 3.4). We then go back to the patient population P_1, ..., P_{50} and also generate a new population with fixed parameters across all patients informed by the DAISY analysis. We fit each population as one large dataset and consider the individual parameter fits that supplement the population fit in Monolix. We compare the effectiveness of these parameter estimates to the estimates from the previous sections where each individual patient was considered its own population (Section 3.5). Finally, we return to the repeated patient population $\mathcal{P}_1$, ..., $\mathcal{P}_{50}$, considering the case where A and I data is observed, and substantially reduce the number of points in the datasets that are provided to Monolix. We compare the errors in the estimates to the case where the datasets were more dense and discuss the implications for practical identifiability (Section 3.6).

3.1 Identifiability analysis observing only the B, A, and I compartments

In Chapter 2, we assumed that all compartments of the LNP model in equations 2.20a – 2.20f were observed as state variables. This of course represents an ideal scenario and in the case of real clinical trials the number of observed state variables is often much smaller than the number of compartments being modelled. As such, we consider here the case where the LNP concentration (L), vaccinated cells (V), and T–cells (T) are no longer observed leaving us only with observations of the plasma B–cells (B), antibody (A), and interleukin (I). This is motivated by the fact that, Korosec et al. [27] only observed data from A and I; hence lacking observations of the linear dynamics upstream. We additionally fix B in this case due to the strong non–linear relationship between B and I that provides a point of comparison as we further reduce the available data in the following sections. Therefore, our analysis here considers an ideal case given the absence of upstream observations leaving only the non–linear dynamics observed.

We first reconsider the algebraic analysis attained from DAISY. In Section 2.2 we proved that when all compartments of the LNP model are observed, the differential algebra system yields global identifiability. Without observations of L, V, and T, the differential algebra system becomes more complicated and so we use DAISY to perform the analysis given the LNP model while only observing B, A, and I. The full system can be seen in Appendix B. As mentioned in Section 1.2, DAISY classifies the model identifiability by assigning prime number values to the model parameters and then solving the resulting differential algebra system. These assigned values p and the corresponding solutions $\widehat{p}$ are shown in Table 3.1.

Parameter	DAISY assigned initial value	Differential algebra solution
μ_{LV}	2	$\widehat{\mu}_{LV} = 2$
μ_{VT}	3	None
μ_{TB}	5	$\widehat{\mu}_{TB} = \frac{5\widehat{\mu}_{TI}}{11}$
μ_{BA}	7	$\widehat{\mu}_{BA} = 7$
μ_{TI}	11	$\widehat{\mu}_{TI} = \frac{11\widehat{\mu}_{TB}}{5}$
γ_V	13	$\widehat{\gamma}_V = 13$ or $\widehat{\gamma}_V = 17$
γ_T	17	$\widehat{\gamma}_T = 13$ or $\widehat{\gamma}_T = 17$
γ_B	19	$\widehat{\gamma}_B = 19$
γ_A	23	$\widehat{\gamma}_A = 23$
γ_I	29	$\widehat{\gamma}_I = 29$
β_{BI}	31	$\widehat{\beta}_{BI} = 31$
β_{IB}	37	$\widehat{\beta}_{IB} = 37$
s_I	41	$\widehat{s}_I = 41$

Table 3.1: DAISY assigned parameter values p and corresponding solutions $\widehat{p}$ of the differential algebra system given the LNP model while only observing B, A, and I as state variables.

Recalling the definitions of identifiability outlined in Section 1.2, we classify each parameter p under the following scheme

- If $p = \widehat{p}$ is a unique differential algebra solution, then p is *globally identifiable*.
- If $p = \widehat{p}$ is part of a finite set of non–unique differential algebra solutions, then p is *locally identifiable*.
- If $\widehat{p}$ is contained within a function of parameters such that $p = \widehat{p}$ is a unique differential algebra solution if one or more other parameters are fixed, then p is *partially identifiable*
- If none of the above are true, i.e. if no solution for $\widehat{p}$ exists, then p is *non–identifiable*

Referring to Table 3.1, we see that many of the parameters have unique solutions matching their assigned values and are hence globally identifiable. For those that are not, we first note that $\widehat{\gamma}_V$ and $\widehat{\gamma}_T$ equal both their assigned values as well as each other's values. So they have non–unique solutions that match their assigned values and hence are locally identifiable. Then we consider μ_{VT} which has no solution in the differential algebra system since it is completely eliminated in the equations and so it is non–identifiable. Finally, we consider $\widehat{\mu}_{TB}$ and $\widehat{\mu}_{TI}$. These parameters do not take on discreet values in the DAISY solutions. Instead, their solutions are defined by the relation

$$\frac{\widehat{\mu}_{TB}}{\widehat{\mu}_{TI}} = \frac{5}{11} = \widehat{k} \tag{3.2}$$

This ratio itself is globally identifiable since

$$k = \frac{\mu_{TB}}{\mu_{TI}} = \frac{5}{11} = \frac{\widehat{\mu}_{TB}}{\widehat{\mu}_{TI}} = \widehat{k}$$

While in general there are no differential algebra solutions such that the parameters individually match their assigned values, i.e. $\widehat{\mu}_{TB} = 5$ and $\widehat{\mu}_{TI} = 11$ explicitly, if one of them is set to its corresponding value, then the relation can be solved to yield the other

$$\widehat{\mu}_{TB} = 5 \Longrightarrow \frac{5}{\widehat{\mu}_{TI}} = \frac{5}{11} \Longrightarrow \widehat{\mu}_{TI} = 11 \tag{3.3}$$

$$\widehat{\mu}_{TI} = 11 \Longrightarrow \frac{\widehat{\mu}_{TB}}{11} = \frac{5}{11} \Longrightarrow \widehat{\mu}_{TB} = 5$$

Therefore fixing one of the parameters from the locally identifiable ratio to its exact value makes the other globally identifiable. And so by virtue of equation 3.3, we say that μ_{TB} and μ_{TI} are partially identifiable. Accordingly, we display the classifications in Table 3.4. Unsurprisingly, by eliminating the upstream compartments as observed state variables, the model is no longer globally identifiable which was the case in Chapter 2. Due to said global identifiability, when we originally estimated the individual parameter values on the patient population $\{P_1, ..., P_{50}\}$ as well as the modified population $\{\mathcal{P}_1, ..., \mathcal{P}_{50}\}$, we could not make any inferences on which parameters we could fix to reduce the errors. In this case however, we can make inferences on which parameters to fix at the beginning given the some are no longer globally identifiable.

Globally identifiable	Locally identifiable	Partially identifiable	Non–identifiable
μ_{LV}	γ_V	μ_{TB}	μ_{VT}
μ_{BA}	γ_T	μ_{TI}	
γ_B			
γ_A			
γ_I			
β_{BI}			
β_{IB}			
s_I			

Table 3.4: Classification of structural identifiability from DAISY solutions for the LNP model while only observing B, A, and I as state variables.

We now return to fitting parameter values on the modified patient population $\{\mathcal{P}_1, ..., \mathcal{P}_{50}\}$ introduced in Section 2.7. We recall that this is simply the patient population $\{P_1, ..., P_{50}\}$ introduced in Section 2.6 with each patient dataset repeated five times and this is done to take better advantage of the population fits from the SAEM algorithm. Like the previous testing, we keep s_I, A_0, and I_0 fixed in all cases. We acknowledge that in this case s_I is globally identifiable however for this testing we choose to remain consistent with Korosec et al. [27] and discuss the case where s_I is unfixed in Section 3.4. Furthermore, given the results of Section 2.9, we also fix β_{IB} in all cases as we suspect that the small size of the parameter makes it harder for the SAEM algorithm to correctly identify it. We then experiment with different fixed parameters sets as before but this time with inferences from Table 3.4. Namely, we experiment with fixing one parameter from the pair (γ_V, γ_T) given the local identifiability and one parameter from the pair (μ_{TB}, μ_{TI}) given the coupling resulting in one of them becoming globally identifiable if the other is fixed. μ_{VT} is fixed in all cases. Just like Section 2.9, we streamline the process by bootstrapping with the patient sample

$$\{\mathcal{P}_{14},\ \mathcal{P}_{16},\ \mathcal{P}_{18},\ \mathcal{P}_{20},\ \mathcal{P}_{30},\ \mathcal{P}_{32},\ \mathcal{P}_{34},\ \mathcal{P}_{35},\ \mathcal{P}_{36},\ \mathcal{P}_{49}\}$$

We recall that this sample represents ten patients with some of the highest errors from the control case in Table 2.23. Using this sample, we test the effectiveness of the Monolix parameter fits under the different fixed parameter sets. Then after identifying fixed sets that yield low errors, we fit the remainder of the patient population. The experimental procedure for each patient is as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set B, A, and I as observed state variables
3. Load the LNP model equations
4. Fix β_{IB}, s_I, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Fix additional parameters with inferences from DAISY based on the results of Table 3.4
6. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distributions
7. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
8. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
9. Run the SAEM algorithm which outputs the population parameter fits

The results of the testing using the bootstrapping sample on different fixed parameter sets are summarised in Table 3.5 where the overall mean errors are displayed with with references to the results table for each case. Like Chapter 2, we begin with a control case where no additional parameters are fixed. Observing these results in Table 3.6, we see a more demonstrated agreement with the structural identifiability results of Table 3.4 compared to what was observed in Chapter 2. The non–identifiable parameter μ_{VT} has the highest mean error being over 200%. The partially identifiable parameters μ_{TB}, and μ_{TI} also have high mean errors with both exceeding 50%. The two locally identifiable parameters γ_V and γ_T have lower but still noticeably high mean errors of 38.50% and 48.78% respectively. Finally, most of the globally identifiable parameters have noticeably low mean errors that do not exceed 5% with the only exceptions being μ_{LV} and γ_I. For

μ_{LV} we suspect that this is due to it being the parameter that is furthest upstream from the observed compartments. This gives an overall mean error of 44.95%. Following the control case, the first two cases that we tested involved fixing μ_{VT} and μ_{TB} while fixing one of the locally identifiable parameters γ_V and γ_T. We left μ_{TI} unfixed due to the coupling. From Table 3.5, we see that the mean errors are noticeably improved over the control with the overall mean errors both being less than 3%. Observing Tables 3.7 and 3.8, we see all of the parameters have mean errors of less than 5% with μ_{LV}, the most upstream parameter from the observed data, being the only exception. We then tested two more cases with μ_{VT} and μ_{TI} fixed while fixing one of γ_V and γ_T. This time μ_{TB} was unfixed due to the coupling. We see a further reduction of the overall mean errors in Table 3.5 with both cases being less than 1%. Then observing Tables 3.9 and 3.10, we see that, barring μ_{LV} in the former, all of the parameters have mean errors less than 2%.

Fixed Parameters	Mean Error	Results Table
None	44.95%	Table 3.6
μ_{VT}, μ_{TB}, γ_V	2.63%	Table 3.7
μ_{VT}, μ_{TB}, γ_T	1.67%	Table 3.8
μ_{VT}, μ_{TI}, γ_V	0.93%	Table 3.9
μ_{VT}, μ_{TI}, γ_T	0.62%	Table 3.10

Table 3.5: Summary of parameter fits observing B, A, and I data attained using the SAEM algorithm on the patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ with different fixed parameter sets informed by the results of Table 3.4. Note that β_{IB}, s_I, A_0, and I_0 are fixed in all cases. The mean errors are calculated by taking the mean of all individual parameter errors across all patients. The specified results tables showcase the full results.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	7.27%	69.85%	40.77%	6.67%	30.95%	38.89%	12.00%	1.82%	2.00%	34.62%	0.82%
16	72.65%	386.08%	65.22%	11.11%	43.81%	153.85%	4.00%	0.00%	0.00%	103.13%	4.32%
18	4.52%	42.79%	29.17%	0.00%	24.34%	10.67%	21.15%	4.00%	2.78%	29.79%	0.00%
20	41.67%	7.60%	42.86%	7.06%	30.38%	10.00%	19.48%	1.82%	0.00%	81.58%	1.07%
30	32.59%	49.45%	178.35%	2.41%	162.03%	63.08%	250.00%	0.00%	0.00%	4.55%	1.23%
32	26.33%	2.58%	21.43%	0.00%	22.87%	22.45%	2.17%	5.56%	2.13%	65.79%	2.19%
34	32.16%	1377.42%	90.96%	0.00%	90.86%	20.21%	65.96%	2.33%	0.00%	28.13%	0.72%
35	25.08%	52.48%	167.61%	1.54%	150.10%	8.33%	10.64%	0.00%	0.00%	12.90%	0.85%
36	0.00%	9.35%	6.06%	3.90%	19.70%	40.91%	85.71%	0.00%	2.63%	0.00%	1.06%
49	13.86%	15.05%	7.14%	12.99%	32.19%	16.67%	16.67%	4.35%	0.00%	96.55%	1.75%
Arithmetic Mean	25.61%	201.26%	64.96%	4.57%	60.72%	38.50%	48.78%	1.99%	0.95%	45.70%	1.40%
Median	25.71%	46.12%	41.81%	3.15%	31.57%	21.33%	18.07%	1.82%	0.00%	32.20%	1.06%

Table 3.6: Percentage errors observing B, A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	9.09%	0.00%	0.00%	3.33%	6.27%	0.00%	0.00%	3.64%	0.00%	1.92%	2.87%
16	17.09%	0.00%	0.00%	5.56%	3.92%	0.00%	0.00%	0.00%	0.00%	6.25%	2.16%
18	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	16.91%	0.00%	0.00%	3.53%	4.27%	0.00%	0.00%	1.82%	0.00%	0.00%	1.71%
30	9.63%	0.00%	0.00%	4.82%	5.70%	0.00%	0.00%	3.23%	0.00%	0.00%	2.46%
32	2.51%	0.00%	0.00%	1.49%	0.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%
34	41.21%	0.00%	0.00%	5.56%	10.63%	0.00%	2.13%	4.65%	0.00%	0.00%	4.80%
35	1.90%	0.00%	0.00%	0.00%	0.44%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
36	2.63%	0.00%	0.00%	1.30%	1.75%	0.00%	0.00%	2.13%	0.00%	0.00%	1.06%
49	2.97%	0.00%	0.00%	0.00%	0.75%	0.00%	0.00%	0.00%	0.00%	3.45%	1.25%
Arithmetic Mean	10.44%			2.56%	3.46%		0.21%	1.55%	0.00%	1.16%	1.68%
Median	6.03%			2.41%	2.83%		0.00%	0.91%	0.00%	0.00%	1.48%

Table 3.7: Percentage errors observing B, A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, and γ_V are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	7.27%	0.00%	0.00%	3.33%	5.84%	0.00%	0.00%	3.64%	0.00%	1.92%	2.46%
16	16.67%	0.00%	0.00%	5.56%	3.44%	0.00%	0.00%	0.00%	0.00%	6.25%	1.54%
18	0.45%	0.00%	0.00%	0.00%	0.17%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	5.64%	0.00%	0.00%	1.18%	0.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.64%
30	2.22%	0.00%	0.00%	1.20%	1.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.49%
32	0.94%	0.00%	0.00%	0.00%	0.36%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	13.07%	0.00%	0.00%	0.00%	2.26%	1.06%	0.00%	0.00%	0.00%	0.00%	0.96%
35	1.27%	0.00%	0.00%	0.00%	0.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
36	7.89%	0.00%	0.00%	2.60%	5.11%	0.00%	0.00%	2.13%	0.00%	2.04%	2.34%
49	5.94%	0.00%	0.00%	1.30%	1.99%	0.00%	0.00%	4.35%	0.00%	3.45%	2.49%
Arithmetic Mean	6.14%			1.52%	2.14%	0.11%		1.01%	0.00%	1.37%	1.11%
Median	5.79%			1.19%	1.63%	0.00%		0.00%	0.00%	0.00%	0.80%

Table 3.8: Percentage errors observing B, A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.41%
16	13.25%	0.00%	3.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	2.47%
18	1.72%	0.00%	4.76%	0.00%	0.00%	0.00%	1.30%	1.82%	0.00%	2.63%	0.85%
20	1.72%	0.00%	4.76%	0.00%	0.00%	0.00%	1.30%	1.82%	0.00%	2.63%	0.85%
30	1.48%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	1.51%	0.00%	2.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	0.96%
35	0.63%	0.00%	1.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%
36	0.00%	0.00%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%
49	0.99%	0.00%	0.00%	1.30%	0.00%	0.00%	0.00%	0.00%	0.00%	6.90%	2.00%
Arithmetic Mean	2.13%		1.84%	0.13%			0.26%	0.36%	0.00%	1.84%	0.86%
Median	1.24%		1.46%	0.00%			0.00%	0.00%	0.00%	1.32%	0.64%

Table 3.9: Percentage errors observing B, A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_V are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	0.00%	0.00%	7.69%	0.00%	0.00%	1.85%	0.00%	1.82%	0.00%	3.85%	1.23%
16	2.56%	0.00%	5.43%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	1.54%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	0.25%	0.00%	2.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%	0.64%
30	2.22%	0.00%	1.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%
34	0.50%	0.00%	2.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%
35	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
36	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
49	1.98%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.45%	1.00%
Arithmetic Mean	0.81%		1.93%	0.00%		0.19%		0.18%	0.00%	1.31%	0.54%
Median	0.37%		0.52%	0.00%		0.00%		0.00%	0.00%	0.00%	0.40%

Table 3.10: Percentage errors observing B, A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

We can state the following conclusions from the results of the bootstrapping fits. Firstly, the effectiveness of DAISY in helping to find less identifiable parameters prior to running the parameter fits is truly showcased here. Unlike Chapter 2 where our approach to fixing parameters was more based on trial and error given the distribution of the errors, the lack of global identifiability in the model here coupled with the results of DAISY immediately suggested a need to fix μ_{VT}, one of μ_{TB} and μ_{TI}, and one of γ_V and γ_T. All of these fixed parameter sets resulted in substantially lower errors compared to the control case and even compared to some of the results of Chapter 2 when all compartments were observed. Of course in all cases, two parameters from the unobserved linear compartments upstream were fixed which naturally reflects the lack of a complete picture of the dynamics. In terms of which particular parameters have a greater impact on error reduction when fixed, we see that the cases where μ_{TI} was fixed have lower mean errors than the cases where μ_{TB} was fixed. We discuss reasons as to why this might be the case in Section 3.3. Similarly, the cases where γ_T was fixed had lower mean errors than the cases where γ_V was fixed. We suspect that this is due to both parameters from T being fixed when fixing γ_T. From this, the lowest mean error of 0.62% involved the fixing of μ_{VT}, μ_{TI}, and γ_T in Table 3.10. We consider this to be an acceptable case of fixing as we have a mean error of less than 1% while only fixing parameters from two compartments, T and I. Naturally, this suggests that in the absence of observations from the upstream linear compartments, a pre–existing understanding of the T–cell and interleukin dynamics on the population level that allows for fixing these parameters can be beneficial towards improving the accuracy of fitting data to the LNP model. This is especially the case for the T–cells since all parameters from T are fixed.

We fit the remainder of the virtual patient population $\{\mathcal{P}_1, \ ..., \ \mathcal{P}_{50}\}$ using the fixed parameter set $\{\mu_{VT}, \ \mu_{TI}, \ \gamma_T\}$ under the same procedure as the bootstrapping sample. The results are shown in Table 3.11. The mean error across all parameters is 1.37% which is lower than the mean error of 1.98% seen in Table 2.29 despite that fact that in this case none of the linear compartments are observed as state variables. Some outliers do exist, most notably $\mathcal{P}_{13}$ which we discussed at the end of Section 2.7 as potentially being the case of a weak immune system. However overall, we consider the errors to be acceptably low especially given that the number of fixed parameters does not exceed those from Section 2.6 and 2.7. To have a point of comparison, since the DAISY results did indicate that β_{IB} is globally identifiable, we repeat fitting the whole population under the same fixed parameter set but with β_{IB} unfixed. The results are shown in Table 3.12. We can see that the mean error for β_{IB} is the largest out of the fitted parameters and in general, the patients for which β_{IB} is fitted poorly also have other parameters fitted poorly with μ_{TB} and μ_{BA} in particular having noticeably higher mean errors. This relationship becomes more clear in the next section where the observation of B removed.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
1	0.58%	0.00%	4.29%	0.00%	0.00%	0.00%	0.00%	1.69%	0.00%	0.00%	1.24%
2	0.24%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3	0.24%	0.00%	2.94%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.67%
4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%
5	1.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	32.43%	0.00%
6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
7	1.12%	0.00%	3.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.45%	0.74%
8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
9	3.06%	0.00%	9.09%	0.00%	0.00%	1.15%	0.00%	0.00%	0.00%	9.52%	1.91%
10	15.38%	0.00%	30.26%	3.03%	0.00%	0.00%	0.00%	0.00%	0.00%	2.56%	1.25%
11	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%
12	1.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.56%	0.56%
13	24.29%	0.00%	7.69%	17.74%	0.00%	11.82%	0.00%	16.00%	0.00%	12.50%	8.78%
14	0.00%	0.00%	7.69%	0.00%	0.00%	1.85%	0.00%	1.82%	0.00%	3.85%	1.23%
15	0.33%	0.00%	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.50%	1.28%
16	2.56%	0.00%	5.43%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	1.54%
17	0.55%	0.00%	1.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.28%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
19	4.68%	0.00%	0.00%	1.04%	0.00%	0.00%	0.00%	0.00%	0.00%	2.56%	0.34%
20	0.25%	0.00%	2.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%	0.64%
21	2.78%	0.00%	8.33%	0.00%	0.00%	1.00%	0.00%	0.00%	0.00%	6.45%	1.52%
22	0.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
23	4.02%	0.00%	3.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.73%	0.74%
24	18.06%	0.00%	8.33%	1.06%	0.00%	0.00%	0.00%	3.64%	0.00%	0.00%	2.73%
25	9.45%	0.00%	4.05%	0.00%	0.00%	1.02%	0.00%	0.00%	0.00%	5.88%	0.59%
26	1.70%	0.00%	3.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.50%	0.54%
27	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.63%
28	0.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
29	4.64%	0.00%	6.90%	0.00%	0.00%	0.00%	0.00%	1.79%	0.00%	0.00%	2.00%
30	2.22%	0.00%	1.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%	0.64%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%
33	1.39%	0.00%	2.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.70%	0.53%
34	0.50%	0.00%	2.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%
35	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
36	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
37	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
38	2.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.22%
39	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	3.16%	0.00%	5.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.96%
41	1.72%	0.00%	1.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.86%	0.25%
42	1.43%	0.00%	16.92%	0.00%	0.00%	1.28%	0.00%	5.26%	0.00%	0.00%	5.19%
43	0.52%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.86%	1.01%
44	8.05%	0.00%	1.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%
45	1.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.71%
46	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.00%	0.27%
47	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.38%	0.51%
48	0.00%	0.00%	1.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.22%
49	1.98%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.45%	1.00%
50	2.90%	0.00%	9.09%	1.22%	0.00%	1.30%	0.00%	0.00%	0.00%	5.56%	1.26%
Arithmetic Mean	2.51%		3.21%	0.48%		0.39%		0.60%	0.00%	2.83%	0.91%
Median	0.61%		1.02%	0.00%		0.00%		0.00%	0.00%	0.00%	0.58%

Table 3.11: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\mathcal{P}_1, \ldots, \mathcal{P}_{50}\}$ generated from the LNP model while only observing B, A, and I data. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}
1	7.49%	0.00%	57.14%	101.85%	0.00%	0.00%	0.00%	3.39%	0.00%	0.00%	4.66%	103.13%
2	2.16%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.86%
3	0.95%	0.00%	8.82%	11.76%	0.00%	0.00%	0.00%	1.89%	0.00%	0.00%	0.67%	11.76%
4	2.24%	0.00%	28.57%	16.67%	0.00%	1.00%	0.00%	3.92%	0.00%	2.44%	3.82%	23.53%
5	17.19%	0.00%	15.00%	9.26%	0.00%	1.19%	0.00%	0.00%	0.00%	16.22%	0.27%	11.11%
6	0.00%	0.00%	7.69%	7.14%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%
7	3.37%	0.00%	8.00%	9.37%	0.00%	0.00%	0.00%	0.00%	0.00%	3.45%	0.37%	11.11%
8	0.00%	0.00%	8.33%	14.86%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%	13.64%
9	1.02%	0.00%	15.45%	11.76%	0.00%	1.15%	0.00%	0.00%	0.00%	9.52%	1.15%	15.15%
10	15.38%	0.00%	56.58%	19.19%	0.00%	9.09%	0.00%	8.16%	0.00%	30.77%	8.11%	26.09%
11	0.00%	0.00%	7.14%	5.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%	6.45%
12	3.82%	0.00%	15.38%	13.51%	0.00%	1.25%	0.00%	0.00%	0.00%	2.56%	0.28%	14.29%
13	30.00%	0.00%	30.00%	74.19%	0.00%	9.09%	0.00%	2.00%	0.00%	5.00%	5.07%	31.25%
14	0.00%	0.00%	85.38%	276.67%	0.00%	1.85%	0.00%	10.91%	0.00%	3.85%	11.48%	284.62%
15	10.96%	0.00%	6.00%	10.00%	0.00%	0.00%	0.00%	3.57%	0.00%	0.00%	1.28%	10.53%
16	3.85%	0.00%	14.13%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.54%	12.90%
17	16.67%	0.00%	33.00%	46.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.28%	46.15%
18	0.45%	0.00%	8.33%	6.19%	0.00%	0.00%	0.00%	0.00%	0.00%	2.13%	0.00%	4.55%
19	0.49%	0.00%	5.48%	4.17%	0.00%	0.00%	0.00%	0.00%	0.00%	2.56%	0.00%	9.09%
20	5.64%	0.00%	17.86%	10.59%	0.00%	1.25%	0.00%	1.82%	0.00%	5.26%	1.49%	9.68%
21	5.56%	0.00%	40.28%	53.33%	0.00%	1.00%	0.00%	0.00%	0.00%	16.13%	0.76%	55.88%
22	0.28%	0.00%	4.48%	6.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	6.06%
23	2.30%	0.00%	46.39%	85.51%	0.00%	2.33%	0.00%	2.86%	4.26%	9.80%	0.37%	86.96%
24	3.23%	0.00%	8.33%	13.83%	0.00%	1.92%	0.00%	7.27%	0.00%	4.44%	5.19%	10.71%
25	13.78%	0.00%	10.81%	21.74%	0.00%	1.02%	0.00%	0.00%	0.00%	2.94%	0.88%	22.22%
26	8.50%	0.00%	16.00%	17.65%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.82%	20.00%
27	8.58%	0.00%	26.15%	35.29%	0.00%	0.00%	0.00%	1.75%	0.00%	0.00%	1.90%	35.71%
28	0.94%	0.00%	35.45%	53.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	52.00%
29	10.60%	0.00%	65.52%	74.07%	0.00%	0.00%	0.00%	5.36%	0.00%	0.00%	7.75%	82.61%
30	0.74%	0.00%	3.09%	3.61%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	2.78%
31	2.65%	0.00%	41.43%	52.58%	0.00%	2.00%	0.00%	0.00%	0.00%	6.45%	1.71%	56.52%
32	0.31%	0.00%	7.14%	7.46%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%	7.89%
33	0.00%	0.00%	58.62%	85.42%	0.00%	3.00%	0.00%	3.13%	0.00%	5.41%	4.47%	91.67%
34	3.02%	0.00%	13.70%	16.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.24%	16.13%
35	0.32%	0.00%	7.04%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%	7.89%
36	0.00%	0.00%	10.61%	11.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%	12.12%
37	0.40%	0.00%	7.14%	3.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.55%
38	5.28%	0.00%	12.82%	16.44%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.89%	16.67%
39	0.79%	0.00%	7.69%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.39%	7.14%
40	2.37%	0.00%	10.59%	13.46%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.64%	13.51%
41	5.17%	0.00%	10.11%	8.70%	0.00%	2.78%	0.00%	2.33%	4.76%	25.71%	0.25%	10.81%
42	1.43%	0.00%	61.54%	66.67%	0.00%	1.28%	0.00%	8.77%	0.00%	3.57%	9.52%	67.74%
43	5.67%	0.00%	15.38%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	5.71%	0.25%	13.33%
44	12.08%	0.00%	8.08%	15.00%	0.00%	2.00%	0.00%	2.94%	0.00%	4.00%	1.20%	14.29%
45	1.03%	0.00%	16.36%	4.08%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	2.13%	8.33%
46	0.00%	0.00%	8.00%	8.33%	0.00%	1.43%	0.00%	0.00%	0.00%	4.00%	0.27%	9.68%
47	0.00%	0.00%	8.33%	9.37%	0.00%	0.00%	0.00%	0.00%	0.00%	7.14%	0.51%	7.69%
48	1.97%	0.00%	9.18%	9.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	13.16%
49	0.99%	0.00%	14.29%	12.99%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.00%	13.89%
50	29.47%	0.00%	34.55%	18.29%	0.00%	5.19%	0.00%	5.56%	0.00%	63.89%	3.15%	23.53%
Arithmetic Mean	4.98%		21.10%	28.21%		1.00%		1.58%	0.18%	4.86%	1.78%	28.94%
Median	2.27%		13.26%	12.38%		0.00%		0.00%	0.00%	0.00%	0.65%	13.42%

Table 3.12: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\mathcal{P}_1, \ldots, \mathcal{P}_{50}\}$ generated from the LNP model while only observing B, A, and I data. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_T are fixed to their exact values while β_{IB} is unfixed and has its results visible. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

3.2 Identifiability analysis observing only the A and I compartments

We now consider the case where only the antibodies (A) and interleukin (I) are observed as state variables which reflects the availability of data in Korosec et al. [27]. We once again look to the algebraic analysis from DAISY. The full system can be seen in Appendix B while the DAISY assigned values p and corresponding differential algebra solutions $\widehat{p}$ are shown in Table 3.13.

Parameter	DAISY assigned initial value	Differential algebra solution
μ_{LV}	2	$\widehat{\mu}_{LV} = 2$
μ_{VT}	3	None
μ_{TB}	5	$\widehat{\mu}_{TB} = \frac{(5)(7)\widehat{\mu}_{TI}}{11\widehat{\mu}_{BA}}$ or $\widehat{\mu}_{TB} = \frac{(5)(37)\widehat{\mu}_{TI}}{11\widehat{\beta}_{IB}}$
μ_{BA}	7	$\widehat{\mu}_{BA} = \frac{(5)(7)\widehat{\mu}_{TI}}{11\widehat{\mu}_{TB}}$
μ_{TI}	11	$\widehat{\mu}_{TI} = \frac{11\widehat{\mu}_{TB}\widehat{\mu}_{BA}}{(5)(7)}$ or $\widehat{\mu}_{TI} = \frac{11\widehat{\mu}_{TB}\widehat{\beta}_{IB}}{(5)(37)}$
γ_V	13	$\widehat{\gamma}_V = 13$ or $\widehat{\gamma}_V = 17$
γ_T	17	$\widehat{\gamma}_T = 13$ or $\widehat{\gamma}_T = 17$
γ_B	19	$\widehat{\gamma}_B = 19$
γ_A	23	$\widehat{\gamma}_A = 23$
γ_I	29	$\widehat{\gamma}_I = 29$
β_{BI}	31	$\widehat{\beta}_{BI} = 31$
β_{IB}	37	$\widehat{\beta}_{IB} = \frac{(5)(37)\widehat{\mu}_{TI}}{11\widehat{\mu}_{TB}}$
s_I	41	$\widehat{s}_I = 41$

Table 3.13: DAISY assigned parameter values and corresponding solutions of the differential algebra system given the LNP model while only observing A, and I as state variables.

These results are very similar to those of Section 3.1 with some small differences. Several parameters have unique solutions matching the assigned values from DAISY and are hence globally identifiable. Out of those that are not, we once again see that γ_V and γ_T take on both their assigned values and each other's values making them locally identifiable. Furthermore, μ_{VT} has no solution at all and is hence non–identifiable. Finally, $\widehat{\mu}_{TB}$, $\widehat{\mu}_{BA}$, $\widehat{\mu}_{TI}$, and $\widehat{\beta}_{IB}$ have their solutions defined by the relations

$$\frac{\widehat{\mu}_{TB}\widehat{\mu}_{BA}}{\widehat{\mu}_{TI}} = \frac{(5)(7)}{11} = \widehat{k}_1, \quad \frac{\widehat{\mu}_{TB}\widehat{\beta}_{IB}}{\widehat{\mu}_{TI}} = \frac{(5)(37)}{11} = \widehat{k}_2 \tag{3.14}$$

Similar to equation 3.2, these two ratios are globally identifiable since

$$k_1 = \frac{\mu_{TB}\mu_{BA}}{\mu_{TI}} = \frac{(5)(7)}{11} = \frac{\widehat{\mu}_{TB}\widehat{\mu}_{BA}}{\widehat{\mu}_{TI}} = \widehat{k}_1$$

$$k_2 = \frac{\mu_{TB}\beta_{IB}}{\mu_{TI}} = \frac{(5)(37)}{11} = \frac{\widehat{\mu}_{TB}\widehat{\beta}_{IB}}{\widehat{\mu}_{TI}} = \widehat{k}_2$$

Like before, while no solutions to the differential algebra system exist such that the parameters explicitly match their DAISY assigned values, they are coupled through the ratios allowing for individual parameters to become globally identifiable when others are fixed. For this, we recall that β_{IB} is kept fixed following Section 2.9 given its small size and the results of Table 3.12 in the previous section where despite it being globally identifiable it was not fitted well. However in this case, its inclusion in the relations here impacts the identifiability of the other parameters when it is fixed. We now consider this by assuming that $\widehat{\beta}_{IB}$ is set to its assigned value of 37. Then second relation of equation 3.14 simplifies and becomes equivalent to the relation in equation 3.2

$$\frac{\widehat{\mu}_{TB}\widehat{\beta}_{IB}}{\widehat{\mu}_{TI}} = \frac{(5)(37)}{11} \Longrightarrow \frac{\widehat{\mu}_{TB}(37)}{\widehat{\mu}_{TI}} = \frac{(5)(37)}{11} \Longrightarrow \frac{\widehat{\mu}_{TB}}{\widehat{\mu}_{TI}} = \frac{5}{11} \tag{3.15}$$

We know that this ratio is globally identifiable and while the individual values of $\widehat{\mu}_{TB}$ and $\widehat{\mu}_{TI}$ are not known, we can consider a general solution $\widehat{\mu}_{TB} = 5m$ and $\widehat{\mu}_{TI} = 11m$ where $m \in \mathbb{R},\ m \neq 0$. From this, the first relation of equation 3.14 simplifies to

$$\frac{\widehat{\mu}_{TB}\widehat{\mu}_{BA}}{\widehat{\mu}_{TI}} = \frac{(5)(7)}{11} \Longrightarrow \frac{(5k)\widehat{\mu}_{BA}}{11k} = \frac{(5)(7)}{11} \Longrightarrow \widehat{\mu}_{BA} = 7 \tag{3.16}$$

Hence, by fixing β_{IB}, μ_{BA} becomes globally identifiable by extension even if μ_{TB} and μ_{TI} are not. Like before, as a result of the relation, μ_{TB} and μ_{TI} are partially identifiable and so knowing one of them automatically yields the other as demonstrated in equation 3.3. Therefore, as a result of the fixing of β_{IB}, we are left with the same classification as Section 3.1. For completeness, we display the classification again in Table 3.17.

Globally identifiable	Locally identifiable	Partially identifiable	Non–identifiable
μ_{LV}	γ_V	μ_{TB}	μ_{VT}
μ_{BA}	γ_T	μ_{TI}	
γ_B			
γ_A			
γ_I			
β_{BI}			
β_{IB}			
s_I			

Table 3.17: Classification of structural identifiability from DAISY solutions for the LNP model while only observing A, and I as state variables and given that β_{IB} is fixed to its exact value.

We now repeat the same parameter fitting experiment with the bootstrapping patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$. The experimental procedure is as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set A, and I as observed state variables
3. Load the LNP model equations
4. Fix β_{IB}, s_I, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Fix additional parameters with inferences from DAISY based on the results of Table 3.17
6. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution
7. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
8. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
9. Run the SAEM algorithm which outputs the population parameter fits

The results of the testing using the bootstrapping sample on different fixed parameter sets are summarised in Table 3.18. Given the identical classification, we test the same fixed parameter sets as Table 3.5. We also once again tested a control case where no additional parameters were fixed. Observing this case in Table 3.19, we see that, unsurprisingly, the errors are higher due to the lack of observations from B compared to before. However we still see some correlation with the DAISY classification. Namely, the non–identifiable priming rate μ_{VT} has the highest mean error of over 160%. The partially identifiable priming rates μ_{TB}, and μ_{TI} and the locally identifiable decay rates γ_V and γ_T also have noticeably high mean errors ranging from 40%–100%. Finally, with the exception of μ_{LV} and γ_I once again, the globally identifiable priming and decay rates all have noticeably low mean errors ranging from 4%–12%. For μ_{LV}, we once again suspect that the higher errors are due to it being the most upstream from the observed compartments. The fits of γ_I are more interesting as the errors, while still being high compared to other globally identifiable parameters are lower on average than those of the case where B was additionally observed in Table 3.6. It is hard to say why this is the case, given that B and I are coupled, but we suspect that

the increased variability from the additional observed compartment may result in a propagation of errors which in this case of Section 3.1 happened to focus on γ_I.

Now observing the cases with the fixed parameter sets, we start with μ_{VT}, μ_{TB}, and one of γ_V and γ_T fixed which can be seen in Tables 3.20 and 3.21. Both of the overall mean errors are less than 3% while all parameters barring μ_{LV} have mean errors of less than 5%. Once again we suspect that the higher errors in μ_{LV} are the result of it being furthest upstream from the observed compartments. Notably, the higher errors seem to be concentrated on particular patients rather than across a given parameter. We then look at the cases where μ_{VT}, μ_{TI}, and one of γ_V and γ_T are fixed which are shown in Tables 3.22 and 3.23. Like before, we see an improvement in the overall mean errors as a result of fixing μ_{TI} instead of μ_{TB} with both being less than 1%. The individual parameter mean errors are also all less than 1% with the exception of μ_{TB} and γ_I in Table 3.23 as a result of slightly higher errors distributed across $\mathcal{P}_{14}$.

Fixed Parameters	Mean Error	Results Table
None	51.91%	Table 3.19
μ_{VT}, μ_{TB}, γ_V	1.98%	Table 3.20
μ_{VT}, μ_{TB}, γ_T	2.58%	Table 3.21
μ_{VT}, μ_{TI}, γ_V	0.14%	Table 3.22
μ_{VT}, μ_{TI}, γ_T	0.58%	Table 3.23

Table 3.18: Summary of parameter fits observing A and I data attained using the SAEM algorithm on the patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ with different fixed parameter sets informed by the results of Table 3.17. Note that β_{IB}, s_I, A_0, and I_0 are fixed in all cases. The mean errors are calculated by taking the mean of all individual parameter errors across all patients. The specified results tables showcase the full results.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	21.82%	33.25%	15.38%	3.33%	7.14%	103.70%	38.00%	0.00%	0.00%	17.31%	0.00%
16	60.68%	421.90%	39.13%	11.11%	59.58%	169.23%	19.00%	15.38%	15.00%	6.25%	10.49%
18	31.67%	343.07%	68.33%	7.22%	70.35%	14.67%	3.85%	0.00%	2.78%	2.13%	0.31%
20	36.03%	132.45%	38.10%	5.88%	43.51%	37.50%	7.79%	1.82%	0.00%	13.16%	1.07%
30	34.81%	36.80%	64.95%	0.00%	63.29%	63.08%	120.00%	6.45%	35.42%	18.18%	2.70%
32	31.66%	33.07%	171.43%	2.99%	125.28%	32.65%	15.22%	11.11%	34.04%	189.47%	8.13%
34	96.43%	581.94%	297.26%	11.11%	123.70%	378.72%	176.60%	11.63%	2.50%	21.88%	15.83%
35	210.60%	0.07%	30.28%	14.00%	17.88%	5.83%	0.64%	2.26%	0.39%	0.65%	1.96%
36	44.74%	17.57%	96.97%	5.19%	115.34%	54.55%	9.52%	2.13%	0.00%	8.16%	1.06%
49	18.81%	4.12%	71.43%	6.49%	46.37%	75.00%	25.00%	13.04%	27.91%	27.59%	7.23%
Arithmetic Mean	58.73%	160.42%	89.33%	6.73%	67.24%	93.49%	41.56%	6.38%	11.80%	30.48%	4.88%
Median	35.42%	35.02%	66.64%	6.19%	61.43%	58.81%	17.11%	4.35%	2.64%	15.23%	2.33%

Table 3.19: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	1.82%	0.00%	0.00%	0.00%	1.70%	0.00%	0.00%	0.00%	0.00%	0.00%	0.82%
16	8.12%	0.00%	0.00%	0.00%	1.48%	0.00%	0.00%	0.00%	0.00%	0.00%	0.62%
18	0.00%	0.00%	0.00%	0.00%	0.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	6.37%	0.00%	0.00%	1.18%	1.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%
30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	19.75%	0.00%	0.00%	2.99%	2.63%	0.00%	8.70%	0.00%	12.77%	39.47%	0.31%
34	13.57%	0.00%	0.00%	0.00%	1.76%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%
35	3.17%	0.00%	0.00%	0.00%	0.87%	0.00%	0.00%	0.00%	0.00%	3.23%	0.21%
36	5.26%	0.00%	0.00%	1.30%	2.99%	0.00%	2.38%	0.00%	0.00%	4.08%	0.64%
49	3.96%	0.00%	0.00%	1.30%	1.99%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
Arithmetic Mean	6.20%			0.68%	1.52%		1.11%	0.00%	1.28%	4.68%	0.40%
Median	4.61%			0.00%	1.61%		0.00%	0.00%	0.00%	0.00%	0.45%

Table 3.20: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, and γ_V are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	32.73%	0.00%	0.00%	13.33%	13.09%	7.41%	0.00%	5.45%	4.00%	17.31%	4.51%
16	5.56%	0.00%	0.00%	0.00%	1.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.31%
18	0.45%	0.00%	0.00%	0.00%	0.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	1.47%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30	1.48%	0.00%	0.00%	1.20%	1.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.49%
32	0.31%	0.00%	0.00%	0.00%	0.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	12.56%	0.00%	0.00%	0.00%	1.18%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%
35	0.63%	0.00%	0.00%	0.00%	0.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	18.42%	0.00%	0.00%	5.19%	19.45%	18.18%	0.00%	4.26%	0.00%	2.04%	3.83%
49	3.96%	0.00%	0.00%	1.30%	2.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.50%
Arithmetic Mean	7.76%			2.10%	3.91%	2.56%		0.97%	0.40%	1.93%	1.01%
Median	2.72%			0.00%	1.17%	0.00%		0.00%	0.00%	0.00%	0.39%

Table 3.21: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
16	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	0.25%	0.00%	1.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
30	0.74%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	1.51%	0.00%	1.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.24%
35	0.63%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
36	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.04%	0.00%
49	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Arithmetic Mean	0.31%		0.26%	0.00%			0.00%	0.00%	0.00%	0.52%	0.07%
Median	0.00%		0.00%	0.00%			0.00%	0.00%	0.00%	0.00%	0.00%

Table 3.22: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_V are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	3.64%	0.00%	7.69%	0.00%	0.00%	5.56%	0.00%	0.00%	4.00%	19.23%	0.82%
16	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	0.25%	0.00%	1.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	1.51%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.24%
35	0.95%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	0.00%	0.00%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
49	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Arithmetic Mean	0.63%		1.04%	0.00%		0.56%		0.00%	0.40%	1.92%	0.13%
Median	0.00%		0.00%	0.00%		0.00%		0.00%	0.00%	0.00%	0.00%

Table 3.23: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Our conclusions following these results are the following. The fixing of μ_{TI} had a greater effect improving the errors compared to the fixing of μ_{TB} which is the same as Section 3.1. This time however, the fixing of γ_V resulted in lower errors compared to the fixing of γ_T. Overall, just like the previous case, the real effectiveness of DAISY in informing our choice of fixed parameters is demonstrated here. Almost all parameter fits using the informed fixed parameter sets have errors less than 5% with many being fitted exactly to their true values. Given that only A and I are observed, we consider this to be an excellent demonstration of finding structural identifiability issues in the LNP model and attaining good fits through fixing the respective parameters. We once again consider the fixed parameter set $\{\mu_{VT},\ \mu_{TI},\ \gamma_T\}$ for testing on the remaining population. Unlike the previous section, this set did not yield the lowest mean error in the bootstrapping. We could choose to fix γ_V instead of γ_T which has slightly lower error in Table 3.18. Both of these cases yielded mean errors of less than 1%. However, we again wish to minimise the number of compartments with fixed parameters and so we accept the slightly higher mean error as a compromise in return for only having two such compartments instead of three. Once again, this requires all parameters from T to be fixed implying the importance of understanding the T–cell dynamics as well as some understanding of the interleukin dynamics given the fixing of μ_{TI}. The results of fitting the full patient population with μ_{VT}, μ_{TI}, and γ_T fixed can be seen in Table 3.24. The overall mean error is 2.47% with a few outliers including once again $\mathcal{P}_{13}$ with low immunity. We consider these errors to be acceptably low.

We briefly comment on the case where only antibodies (A) are observed as a state variable. In this case the computation of the differential algebra system became too difficult for DAISY to accomplish on the hardware that was used. Given this, in a similar manner to Chapter 2, we infer parameters to fix based on the results of the previous sections. This analysis is much shorter than what we have demonstrated so far in this chapter due to the lack of differential algebra solutions and does not warrant its own section. Now given A as the only observed state variable, we once again consider the virtual patient sample $\{\mathcal{P}_{14},\ \mathcal{P}_{16},\ \mathcal{P}_{18},\ \mathcal{P}_{20},\ \mathcal{P}_{30},\ \mathcal{P}_{32},\ \mathcal{P}_{34},\ \mathcal{P}_{35},\ \mathcal{P}_{36},\ \mathcal{P}_{49}\}$. As before, we have a control case where no additional parameters are fixed. For a baseline fixed parameter set, we consider $\{\mu_{VT},\ \mu_{TI},\ \gamma_T\}$ which is what we have used to test the entire patient population in both sections thus far.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
1	0.58%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2	3.61%	0.00%	0.00%	0.00%	0.00%	1.72%	0.00%	0.00%	4.08%	38.46%	0.43%
3	0.71%	0.00%	1.47%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%
5	11.72%	0.00%	10.00%	0.00%	0.00%	1.19%	0.00%	0.00%	0.00%	29.73%	0.27%
6	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
7	0.56%	0.00%	1.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.45%	0.00%
8	1.67%	0.00%	16.67%	2.70%	0.00%	0.00%	0.00%	3.03%	0.00%	1.85%	4.02%
9	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.38%
10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.42%
11	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.38%
12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.28%
13	404.29%	0.00%	46.15%	25.81%	0.00%	18.18%	0.00%	30.00%	2.44%	12.50%	36.49%
14	3.64%	0.00%	7.69%	0.00%	0.00%	5.56%	0.00%	0.00%	4.00%	19.23%	0.82%
15	0.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
16	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
17	0.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
19	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20	0.25%	0.00%	1.19%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.21%
21	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%	0.00%
22	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
23	5.75%	0.00%	13.40%	1.45%	0.00%	11.63%	0.00%	14.29%	31.91%	35.29%	2.21%
24	30.97%	0.00%	0.00%	1.06%	0.00%	0.00%	0.00%	3.64%	0.00%	2.22%	2.46%
25	22.44%	0.00%	1.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.94%	0.88%
26	6.46%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.18%	0.00%	10.00%	1.09%
27	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.42%
28	0.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.45%
29	0.33%	0.00%	2.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%
30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
33	0.28%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
34	1.51%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.24%
35	0.95%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
36	0.00%	0.00%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
37	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
38	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
39	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
40	0.79%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
41	0.00%	0.00%	1.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
42	1.43%	0.00%	1.54%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%
43	3.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
44	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
45	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
46	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
47	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
48	0.85%	0.00%	1.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
49	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
50	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.32%
Arithmetic Mean	10.07%		2.13%	0.62%		0.77%		1.04%	0.85%	3.18%	1.06%
Median	0.25%		0.00%	0.00%		0.00%		0.00%	0.00%	0.00%	0.00%

Table 3.24: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\mathcal{P}_1, \ldots, \mathcal{P}_{50}\}$ generated from the LNP model while only observing A and I data. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

The experimental procedure is as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set A as the observed state variable
3. Load the LNP model equations
4. Fix β_{IB}, s_I, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Fix additional parameters with inferences from DAISY based on the results of Tables 3.4 and 3.17
6. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution
7. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
8. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
9. Run the SAEM algorithm which outputs the population parameter fits

The results of testing on the sample are summarised in Table 3.25. Unsurprisingly, the mean error is noticeably higher in the control case and even with the majority of parameters fixed, it fails to fall below 10%. Since A is observed, the two associated parameters μ_{BA} and γ_A are not fixed except in the last case of Table 3.29 where we fixed μ_{BA}. In general, these two have the lowest mean errors in all cases which which aligns with the observation. We consider the fixed parameter set $\{\mu_{VT},\ \mu_{TB},\ \mu_{TI},\ \gamma_T,\ \gamma_I\}$ from Table 3.28 since we would prefer not to fix μ_{BA} given that we still observe A as a state variable. We fit the remainder of the patient population under this set and the results can be seen in Table 3.30. Once again the lowest mean errors are in μ_{BA} and γ_A from the observed compartment while we see that the higher errors are more even distributed across μ_{LV}, γ_V, and γ_B while β_{BI} has a more varied distribution of low and high errors. One error that stands out is μ_{LV} from $\mathcal{P}_{19}$ which exceeds $10,000\%$ and is the result of the SAEM algorithm diverging to a large value. Including this result, the mean error for the whole population is 64.14% and if we consider it as an outlier and eliminate it, the mean error is 29.04%. Since these results are inconclusive as we lack a classification of identifiability to compare to and given that Korosec et al. [27] had data from I available, for the remainder of this chapter, we consider A and I as observed state variables.

Fixed Parameters	Mean Error	Results Table
None	197.30%	Table 3.26
μ_{VT}, μ_{TB}, μ_{TI}, γ_T	22.22%	Table 3.27
μ_{VT}, μ_{TB}, μ_{TI}, γ_T, γ_I	17.49%	Table 3.28
μ_{VT}, μ_{TB}, μ_{BA}, μ_{TI}, γ_T, γ_I	15.25%	Table 3.29

Table 3.25: Summary of parameter fits observing only A data attained using the SAEM algorithm on the patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ with different fixed parameter sets. Note that β_{IB}, s_I, A_0, and I_0 are fixed in all cases. The mean errors are calculated by taking the mean of all individual parameter errors across all patients. The specified results tables showcase the full results.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	578.18%	28.18%	40.77%	33.33%	45.90%	70.37%	25.00%	56.36%	8.00%	61.54%	5.74%
16	266.67%	92.19%	89.78%	47.78%	29.10%	7.69%	20.00%	23.08%	7.50%	68.75%	17.90%
18	2462.44%	2.11%	51.67%	9.28%	35.22%	14.67%	86.54%	48.00%	13.89%	261.70%	39.94%
20	288.48%	2.96%	44.05%	12.94%	44.67%	50.00%	7.79%	41.82%	8.16%	36.84%	39.87%
30	179.26%	55.07%	198.97%	36.14%	634.81%	26.92%	30.00%	3.23%	39.58%	31.82%	32.92%
32	7205.96%	10.87%	39.29%	35.82%	40.50%	22.45%	95.65%	62.22%	17.02%	55.26%	22.50%
34	76.38%	1440.00%	71.23%	5.56%	31.03%	2612.77%	19.15%	44.19%	5.00%	159.38%	28.78%
35	540.32%	3.84%	39.44%	10.77%	23.40%	0.00%	27.66%	35.48%	23.53%	83.87%	16.38%
36	44.74%	45.69%	31.82%	5.19%	85.66%	9.09%	33.33%	34.04%	13.16%	38.78%	14.47%
49	285.15%	28.00%	37.86%	41.56%	97.33%	31.67%	41.67%	56.52%	11.63%	3244.83%	42.89%
Arithmetic Mean	1192.76%	170.89%	64.49%	23.84%	106.76%	284.56%	38.68%	40.49%	14.75%	404.28%	26.14%
Median	286.81%	28.09%	42.41%	23.14%	42.59%	24.69%	28.83%	43.00%	12.39%	65.14%	25.64%

Table 3.26: Percentage errors observing only A data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Parameter / Patient ID	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	21.82%	0.00%	0.00%	10.00%	0.00%	9.26%	0.00%	1.82%	0.00%	0.00%	25.41%
16	75.64%	0.00%	0.00%	16.67%	0.00%	0.00%	0.00%	38.46%	0.00%	50.00%	67.28%
18	5.88%	0.00%	0.00%	4.12%	0.00%	9.33%	0.00%	24.00%	19.44%	99.96%	3.41%
20	39.95%	0.00%	0.00%	8.24%	0.00%	37.50%	0.00%	32.73%	16.33%	99.98%	11.73%
30	3.70%	0.00%	0.00%	0.00%	0.00%	23.08%	0.00%	29.03%	14.58%	22.73%	0.49%
32	8.78%	0.00%	0.00%	2.99%	0.00%	12.24%	0.00%	11.11%	4.26%	0.00%	7.50%
34	20.60%	0.00%	0.00%	22.22%	0.00%	38.30%	0.00%	34.88%	10.00%	98.13%	11.03%
35	3.81%	0.00%	0.00%	1.54%	0.00%	41.67%	0.00%	32.26%	1.96%	6.45%	0.64%
36	21.05%	0.00%	0.00%	6.49%	0.00%	27.27%	0.00%	38.30%	15.79%	99.92%	6.17%
49	5.94%	0.00%	0.00%	1.30%	0.00%	16.67%	0.00%	17.39%	18.60%	71.03%	16.46%
Arithmetic Mean	20.72%			7.36%		21.53%		26.00%	10.10%	54.82%	15.01%
Median	14.69%			5.31%		19.87%		30.65%	12.29%	60.52%	9.27%

Table 3.27: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, μ_{TI}, and γ_T are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Parameter / Patient ID	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	16.36%	0.00%	0.00%	3.33%	0.00%	29.63%	0.00%	16.36%	2.00%	0.00%	11.89%
16	73.50%	0.00%	0.00%	16.67%	0.00%	7.69%	0.00%	23.08%	0.00%	0.00%	54.94%
18	1.81%	0.00%	0.00%	4.12%	0.00%	46.67%	0.00%	36.00%	5.56%	0.00%	1.24%
20	15.93%	0.00%	0.00%	1.18%	0.00%	21.25%	0.00%	12.73%	0.00%	0.00%	3.62%
30	5.19%	0.00%	0.00%	3.61%	0.00%	69.23%	0.00%	45.16%	6.25%	0.00%	2.46%
32	52.35%	0.00%	0.00%	19.40%	0.00%	42.86%	0.00%	27.78%	2.13%	0.00%	29.38%
34	30.65%	0.00%	0.00%	11.11%	0.00%	17.02%	0.00%	13.95%	0.00%	0.00%	11.75%
35	8.89%	0.00%	0.00%	0.00%	0.00%	41.67%	0.00%	25.81%	1.96%	0.00%	2.13%
36	7.89%	0.00%	0.00%	3.90%	0.00%	36.36%	0.00%	23.40%	2.63%	0.00%	9.79%
49	4.95%	0.00%	0.00%	2.60%	0.00%	41.67%	0.00%	30.43%	6.98%	0.00%	2.74%
Arithmetic Mean	21.75%			6.59%		35.40%		25.47%	2.75%		12.99%
Median	12.41%			3.76%		39.02%		24.61%	2.06%		6.71%

Table 3.28: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, μ_{TI}, γ_T, and γ_I are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
14	21.82%	0.00%	0.00%	0.00%	0.00%	42.59%	0.00%	20.00%	0.00%	0.00%	12.70%
16	50.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.38%	0.00%	0.00%	20.68%
18	2.26%	0.00%	0.00%	0.00%	0.00%	18.67%	0.00%	16.00%	2.78%	0.00%	0.31%
20	18.87%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	23.64%	6.12%	0.00%	4.48%
30	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
32	10.66%	0.00%	0.00%	0.00%	0.00%	32.65%	0.00%	22.22%	0.00%	0.00%	1.87%
34	90.45%	0.00%	0.00%	0.00%	0.00%	6.38%	0.00%	11.63%	0.00%	0.00%	2.88%
35	11.11%	0.00%	0.00%	0.00%	0.00%	50.00%	0.00%	32.26%	1.96%	0.00%	2.34%
36	0.00%	0.00%	0.00%	0.00%	0.00%	36.36%	0.00%	23.40%	0.00%	0.00%	2.98%
49	21.78%	0.00%	0.00%	0.00%	0.00%	58.33%	0.00%	26.09%	2.33%	0.00%	13.72%
Arithmetic Mean	22.70%					27.00%		19.06%	1.32%		6.20%
Median	14.99%					28.83%		21.11%	0.00%		2.93%

Table 3.29: Percentage errors observing A, and I data from fitting parameters simultaneously using the SAEM algorithm on the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ generated from the LNP model. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, μ_{BA}, μ_{TI}, γ_T, and γ_I are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

The results of Table 3.30 provide a great demonstration on the power of observing I as a state variable alongside A compared to just observing A. We go from having an overall mean error of just 2.47% when both A and I are observed in Table 3.24 to an overall mean error of 64.14% (29.04% without the outlier) by removing the I observation despite fixing two additional parameters in μ_{TB} and γ_I. While acknowledging that our choice of fixed parameters was less informed by DAISY given the difficulty of computing the differential algebra system, and given that both compartments were observed by Korosec et al. [27], we conclude this section by proposing that the collection of interleukin (I) data should be considered an essential pairing with collecting antibody (A) data.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}
1	31.41%	0.00%	0.00%	12.96%	0.00%	33.33%	0.00%	27.12%	0.00%	0.00%	7.76%
2	88.22%	0.00%	0.00%	31.00%	0.00%	43.10%	0.00%	275.00%	26.53%	0.00%	120.17%
3	1.66%	0.00%	0.00%	5.88%	0.00%	84.62%	0.00%	41.51%	4.00%	0.00%	3.36%
4	21.08%	0.00%	0.00%	8.33%	0.00%	30.00%	0.00%	25.49%	0.00%	0.00%	1.91%
5	10.94%	0.00%	0.00%	3.70%	0.00%	7.14%	0.00%	25.00%	10.26%	0.00%	5.31%
6	17.42%	0.00%	0.00%	3.57%	0.00%	30.67%	0.00%	25.53%	6.67%	0.00%	0.79%
7	91.57%	0.00%	0.00%	21.88%	0.00%	28.57%	0.00%	35.29%	0.00%	0.00%	104.81%
8	56.67%	0.00%	0.00%	48.65%	0.00%	9.09%	0.00%	100.00%	0.00%	0.00%	55.78%
9	40.82%	0.00%	0.00%	11.76%	0.00%	49.43%	0.00%	23.08%	5.00%	0.00%	31.68%
10	7.69%	0.00%	0.00%	2.02%	0.00%	154.55%	0.00%	55.10%	0.00%	0.00%	5.82%
11	33.33%	0.00%	0.00%	15.00%	0.00%	85.71%	0.00%	48.39%	1.96%	0.00%	16.35%
12	0.00%	0.00%	0.00%	2.70%	0.00%	37.50%	0.00%	24.07%	2.13%	0.00%	1.96%
13	68.57%	0.00%	0.00%	125.81%	0.00%	27.27%	0.00%	28.00%	0.00%	0.00%	41.22%
14	16.36%	0.00%	0.00%	3.33%	0.00%	29.63%	0.00%	16.36%	2.00%	0.00%	11.89%
15	7.31%	0.00%	0.00%	8.33%	0.00%	33.33%	0.00%	25.00%	2.63%	0.00%	0.00%
16	73.50%	0.00%	0.00%	16.67%	0.00%	7.69%	0.00%	23.08%	0.00%	0.00%	54.94%
17	9.84%	0.00%	0.00%	0.00%	0.00%	1.54%	0.00%	3.39%	0.00%	0.00%	0.56%
18	1.81%	0.00%	0.00%	4.12%	0.00%	46.67%	0.00%	36.00%	5.56%	0.00%	1.24%
19	10560.34%	0.00%	0.00%	25.00%	0.00%	15.38%	0.00%	22.22%	0.00%	0.00%	10.88%
20	15.93%	0.00%	0.00%	1.18%	0.00%	21.25%	0.00%	12.73%	0.00%	0.00%	3.62%
21	22.22%	0.00%	0.00%	0.00%	0.00%	10.00%	0.00%	3.16%	0.00%	0.00%	15.66%
22	18.84%	0.00%	0.00%	10.77%	35.88%	44.83%	86.05%	16.67%	0.00%	19.57%	13.73%
23	13.22%	0.00%	0.00%	1.45%	0.00%	6.98%	0.00%	5.71%	6.38%	0.00%	6.99%
24	2.58%	0.00%	0.00%	3.19%	0.00%	1.92%	0.00%	5.45%	0.00%	0.00%	1.91%
25	25.20%	0.00%	0.00%	4.35%	0.00%	2.04%	0.00%	6.45%	0.00%	0.00%	5.01%
26	103.74%	0.00%	0.00%	0.00%	0.00%	33.08%	0.00%	64.71%	0.00%	0.00%	10.63%
27	14.55%	0.00%	0.00%	14.71%	0.00%	650.00%	0.00%	78.95%	12.24%	0.00%	34.25%
28	56.29%	0.00%	0.00%	13.04%	0.00%	30.77%	0.00%	15.38%	0.00%	0.00%	22.73%
29	62.91%	0.00%	0.00%	22.22%	0.00%	45.45%	0.00%	32.14%	0.00%	0.00%	15.75%
30	5.19%	0.00%	0.00%	3.61%	0.00%	69.23%	0.00%	45.16%	6.25%	0.00%	2.46%
31	9.27%	0.00%	0.00%	7.22%	0.00%	50.00%	0.00%	43.33%	22.50%	0.00%	2.13%
32	52.35%	0.00%	0.00%	19.40%	0.00%	42.86%	0.00%	27.78%	2.13%	0.00%	29.38%
33	4.71%	0.00%	0.00%	4.17%	0.00%	60.00%	0.00%	34.38%	7.50%	0.00%	2.11%
34	30.65%	0.00%	0.00%	11.11%	0.00%	17.02%	0.00%	13.95%	0.00%	0.00%	11.75%
35	8.89%	0.00%	0.00%	0.00%	0.00%	41.67%	0.00%	25.81%	1.96%	0.00%	2.13%
36	7.89%	0.00%	0.00%	3.90%	0.00%	36.36%	0.00%	23.40%	2.63%	0.00%	9.79%
37	0.40%	0.00%	0.00%	0.00%	0.00%	6.85%	0.00%	6.59%	0.00%	0.00%	0.42%
38	44.22%	0.00%	0.00%	27.40%	0.00%	29.69%	0.00%	28.95%	0.00%	0.00%	19.82%
39	50.20%	0.00%	0.00%	20.51%	0.00%	30.95%	0.00%	22.86%	0.00%	0.00%	22.09%
40	72.73%	0.00%	0.00%	23.08%	0.00%	100.00%	0.00%	28.57%	0.00%	0.00%	69.43%
41	10.34%	0.00%	0.00%	4.35%	0.00%	19.44%	0.00%	37.21%	4.76%	0.00%	9.23%
42	62.86%	0.00%	0.00%	83.33%	0.00%	41.03%	0.00%	129.82%	0.00%	0.00%	94.81%
43	20.62%	0.00%	0.00%	11.11%	0.00%	20.75%	0.00%	15.00%	4.26%	0.00%	15.19%
44	72.48%	0.00%	0.00%	12.50%	0.00%	110.00%	0.00%	26.47%	0.00%	0.00%	34.26%
45	15.90%	0.00%	0.00%	4.08%	0.00%	42.86%	0.00%	30.00%	5.00%	15.63%	0.95%
46	41.82%	0.00%	0.00%	15.63%	0.00%	37.14%	0.00%	71.43%	0.00%	0.00%	6.52%
47	6.01%	0.00%	0.00%	9.37%	0.00%	33.33%	0.00%	23.08%	5.13%	0.00%	1.28%
48	76.90%	0.00%	0.00%	20.91%	0.00%	779.31%	0.00%	76.92%	5.13%	0.00%	160.17%
49	4.95%	0.00%	0.00%	2.60%	0.00%	41.67%	0.00%	30.43%	6.98%	0.00%	2.74%
50	25.60%	0.00%	0.00%	6.10%	0.00%	55.84%	0.00%	33.33%	11.11%	0.00%	15.77%
Arithmetic Mean	241.96%			14.24%		65.35%		37.51%	3.41%		22.38%
Median	20.85%			8.33%		33.33%		26.79%	0.98%		10.21%

Table 3.30: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\mathcal{P}_1, ..., \mathcal{P}_{50}\}$ generated from the LNP model while only observing A data. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TB}, μ_{TI}, γ_T, and γ_I are fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%. Note that β_{IB}, s_I, A_0, and I_0 are not fitted since they are fixed to their exact values in all cases.

3.3 Parameter fixing contributing to identifiability

In Sections 3.1 and 3.2, we observed different cases of parameters becoming more identifiable when other parameters are fixed. Unlike the cases in Chapter 2 where the model in DAISY was globally identifiable given that all compartments were observed, we can instead infer more about why certain parameters are fitted poorly in the control cases and improve when fixing parameters given the DAISY identifiability classifications of Tables 3.4 and 3.17. We first look at the choice of fixing of either γ_V or γ_T; this being the simplest case. As they both take on their assigned values and each other's values in the DAISY solutions of Tables 3.1 and 3.13, fixing one of them to the assigned value immediately makes the other globally identifiable. This is unsurprisingly reflected by the results of Tables 3.7 – 3.10 and Tables 3.20 – 3.23 where both γ_V and γ_T have mean errors of less than 3% when the other is fixed. We now consider the locally identifiable ratio μ_{TB}/μ_{TI} from the relation in equation 3.2. From the control cases of Tables 3.6 and 3.19, we observed high mean errors for μ_{TB} and μ_{TI} which is in alignment with the fact the the parameters are individually non–identifiable but coupled through the ratio. So when comparing this to Tables 3.7 – 3.10 and Tables 3.20 – 3.23, we see that whenever one of μ_{TB} or μ_{TI} is fixed, the other parameter has a mean error of less than 5%. This verifies the result of equation 3.3; by setting one of the two parameters to its DAISY assigned value, the relation can be solved to attain the other.

The case is similar for equation 3.14. Given the fixing of β_{IB}, the relations simplify to yield μ_{BA} as globally identifiable while leaving the locally identifiable ratio μ_{TB}/μ_{TI} which is what results in the same classification. This is supported by the results of the control case in Table 3.19 where the mean error of μ_{BA} is only 6.73% which is much lower than the mean errors of μ_{TB} and μ_{TI} despite it being similarly coupled in the DAISY solutions. We can further verify this by considering the fitted values of the ratio μ_{TB}/μ_{TI}. While the errors for μ_{TB} and μ_{TI} are high, we suspect that the ratio is identified even if the individual parameters are not since neither are fixed. If this is the case, then μ_{BA} becomes globally identifiable and therefore would have low errors. Table 3.31 shows a comparison of the exact values, fitted values, and errors of μ_{TB}, μ_{TI}, and μ_{TB}/μ_{TI}. Comparing the errors of the ratio to those of the individual fits, we can see that in general, low errors in the ratio correspond to low errors in μ_{BA}. On the other hand, high ratio errors such as those in $\mathcal{P}_{16}$ and $\mathcal{P}_{34}$ correspond to slightly higher errors in μ_{BA}. Most notably, aside from $\mathcal{P}_{14}$ and $\mathcal{P}_{35}$, none of the patients in the sample have errors in μ_{TB} and μ_{TI} lower than 20%. Meanwhile, all errors in μ_{BA} remain within 12%. While the sample size is small, this serves as an indicator that some parameters can be fitted with low errors even if other parameters in the relations attained from the differential algebra system are not. We highlight this as an appropriate demonstration on the effects of fixing parameters in improving structural identifiability allowed by the differential algebra analysis.

Patient	μ_{TB} exact	μ_{TI} exact	μ_{BA} exact	μ_{TB}/μ_{TI} exact	μ_{TB} fitted	μ_{TI} fitted	μ_{BA} fitted	μ_{TB}/μ_{TI} fitted	μ_{TB} error	μ_{TI} error	μ_{BA} error	μ_{TB}/μ_{TI} error
$\mathcal{P}_{14}$	0.130	25.36	0.300	0.00513	0.110	23.55	0.290	0.00467	15.38%	7.14%	3.33%	8.97%
$\mathcal{P}_{16}$	0.092	18.90	0.180	0.00487	0.056	7.64	0.200	0.00732	39.13%	59.58%	11.11%	50.31%
$\mathcal{P}_{18}$	0.120	23.34	0.970	0.00514	0.038	6.92	0.900	0.00549	68.33%	70.35%	7.22%	6.81%
$\mathcal{P}_{20}$	0.084	24.13	0.850	0.00348	0.052	13.63	0.800	0.00382	38.10%	43.51%	5.88%	9.77%
$\mathcal{P}_{30}$	0.097	1.58	0.830	0.06140	0.160	2.58	0.830	0.06200	64.95%	63.29%	0.00%	0.98%
$\mathcal{P}_{32}$	0.140	22.47	0.670	0.00623	0.380	50.62	0.650	0.00751	171.43%	125.28%	2.99%	20.55%
$\mathcal{P}_{34}$	0.073	22.11	0.180	0.00330	0.290	49.46	0.160	0.00586	297.26%	123.70%	11.11%	77.58%
$\mathcal{P}_{35}$	0.071	20.64	0.650	0.00344	0.058	16.57	0.640	0.00350	18.31%	19.72%	1.54%	1.74%
$\mathcal{P}_{36}$	0.066	8.02	0.770	0.00823	0.130	17.27	0.730	0.00753	96.97%	115.34%	5.19%	8.51%
$\mathcal{P}_{49}$	0.140	14.60	0.077	0.00959	0.240	21.37	0.082	0.01120	71.43%	46.37%	6.49%	16.79%

Table 3.31: Comparison of the exact MATLAB values, Monolix fitted values, and errors for μ_{TB}, μ_{TI}, μ_{BA}, and the ratio μ_{TB}/μ_{TI} given the virtual patient sample $\{\mathcal{P}_{14}, \mathcal{P}_{16}, \mathcal{P}_{18}, \mathcal{P}_{20}, \mathcal{P}_{30}, \mathcal{P}_{32}, \mathcal{P}_{34}, \mathcal{P}_{35}, \mathcal{P}_{36}, \mathcal{P}_{49}\}$ in the control case from Table 3.19. Note that the ratio vales are rounded to 3 significant figures while the ratio errors are rounded to 2 decimal places.

Referring back to Section 3.1, we mentioned that the fixing of μ_{TI} improved the errors of the parameter fits more than the fixing of μ_{TB}. Here we provide a more detailed discussion as to why this might be the case. From Table 3.31, we note that the ratio μ_{TB}/μ_{TI} is quite small at $\mathcal{O}(10^{-3})$. This is not just the case with the bootstrapping sample as from Table 2.5, we calculate the ratio of the expected values to be

$$\frac{E(\mu_{TB})}{E(\mu_{TI})} = \frac{0.103}{14.18925} \approx 0.00726 \tag{3.32}$$

So in general, we can in fact expect most virtual patients to have the ratio μ_{VT}/μ_{TI} be roughly $\mathcal{O}(10^{-3})$. If we consider this ratio as a small parameter $\varepsilon \sim \mathcal{O}(10^{-3})$, then we have

$$\frac{\mu_{TB}}{\mu_{TI}} = \varepsilon \Longrightarrow \mu_{TB} = \varepsilon\mu_{TI},\ \mu_{TI} = \frac{1}{\varepsilon}\mu_{TB} \tag{3.33}$$

If μ_{TI} is fixed, then we identify μ_{TB} from the term $\varepsilon\mu_{TI}$. Given that $E(\mu_{TI}) \sim \mathcal{O}(10^2)$, we have $\varepsilon\mu_{TI} \sim \mathcal{O}(10^{-1})$ which is relatively small. Conversely, if μ_{TB} is fixed, then we identify μ_{TI} from μ_{TB}/ε. Given that $E(\mu_{TB}) \sim \mathcal{O}(10^{-1})$, we have $\mu_{TB}/\varepsilon \sim \mathcal{O}(10^2)$ which is large. So since $\varepsilon\mu_{TI}$ is much smaller than μ_{TB}/ε, but still not so small that it compares to parameters like β_{IB} which we believe are harder to fit under the given conditions of SAEM, we consider it to be a more sensibly scaled term and this adds credence to the fact that fixing μ_{TI} yielded better errors than fixing μ_{TB}.

3.4 Variability of s_I

Throughout our testing on the LNP model, we have maintained the fixing of the interleukin saturation $s_I = 1000$ for all virtual patients as was the case in Korosec et al. [27]. However, in both

Tables 3.4 and 3.17, our classification of identifiability from the results of the DAISY differential algebra systems implied that s_I is globally identifiable. In this section, we consider the case where s_I is variable so that we may compare to the fixed case and verify if the fixing was necessary. For this, we return to setting both A and I as observed state variables and consider the fixed parameter set $\{\mu_{VT},\ \mu_{TI},\ \gamma_T\}$ for which we fitted the entire patient population $\{\mathcal{P}_1,\ ...,\ \mathcal{P}_{50}\}$ in Table 3.24. Combined with the parameters which were always fixed, excluding s_I itself, the full fixed parameter set is $\{\mu_{VT},\ \mu_{TI},\ \gamma_T,\ \beta_{IB},\ A_0,\ I_0\}$. Therefore for fitting the full population, the experimental procedure is as follows

1. Load the solution dataset $\mathcal{U}_i$
2. Set A, and I as observed state variables
3. Load the LNP model equations
4. Fix μ_{VT}, μ_{TI}, γ_T, β_{IB}, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution
6. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
7. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
8. Run the SAEM algorithm which outputs the population parameter fits

The results of fitting the patient population are shown in Table 3.34. We see that the errors are noticeably higher with the mean error here being 18.78% compared to the 2.47% of Table 3.24. In fact, s_I has the highest mean and median errors out of the fitted parameters despite the fact that it should be globally identifiable based on the results of Table 3.17. Notably, barring a few outliers such as $\mathcal{P}_{13}$ with weak immunity, in general higher errors in s_I correspond to higher errors in β_{BI} and vice–versa. This is similar to the errors of μ_{TB} and μ_{TI} in the A and I data control case from Table 3.19 which we then showed in Table 3.31 was the result of the ratio being locally identifiable. We can apply this to β_{BI} and s_I. Recalling equation 2.20d, the term in question is

$$\beta_{BI}\left(\frac{I}{I+s_I}\right)B$$

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	1.15%	0.00%	2.86%	1.85%	0.00%	0.00%	0.00%	8.47%	0.00%	0.00%	25.16%	0.00%	43.47%	0.00%	0.00%
2	67.07%	0.00%	41.54%	0.00%	0.00%	10.34%	0.00%	8.33%	18.37%	69.23%	62.47%	0.00%	65.82%	0.00%	0.00%
3	0.95%	0.00%	0.00%	1.96%	0.00%	0.00%	0.00%	1.89%	0.00%	0.00%	2.35%	0.00%	5.91%	0.00%	0.00%
4	2.24%	0.00%	28.57%	0.00%	0.00%	0.00%	0.00%	23.53%	0.00%	0.00%	79.77%	0.00%	173.65%	0.00%	0.00%
5	20.31%	0.00%	15.00%	1.85%	0.00%	3.57%	0.00%	0.00%	2.56%	18.92%	17.51%	0.00%	19.78%	0.00%	0.00%
6	0.28%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	1.06%	0.00%	0.00%	6.82%	0.00%	9.18%	0.00%	0.00%
7	0.00%	0.00%	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.85%	0.00%	2.09%	0.00%	0.00%
8	1.67%	0.00%	30.00%	1.35%	0.00%	0.00%	0.00%	6.06%	0.00%	0.00%	19.10%	0.00%	29.86%	0.00%	0.00%
9	3.06%	0.00%	10.91%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.31%	0.00%	13.60%	0.00%	0.00%
10	0.00%	0.00%	9.21%	0.00%	0.00%	0.00%	0.00%	8.16%	0.00%	0.00%	4.57%	0.00%	15.24%	0.00%	0.00%
11	0.00%	0.00%	21.43%	5.00%	0.00%	0.00%	0.00%	9.68%	0.00%	0.00%	54.37%	0.00%	82.92%	0.00%	0.00%
12	3.44%	0.00%	31.54%	2.70%	0.00%	1.25%	0.00%	11.11%	0.00%	0.00%	28.01%	0.00%	49.39%	0.00%	0.00%
13	350.00%	0.00%	97.62%	3716.13%	0.00%	9.09%	0.00%	134.00%	0.00%	17.50%	147.30%	0.00%	0.07%	0.00%	0.00%
14	9.09%	0.00%	23.08%	0.00%	0.00%	9.26%	0.00%	21.82%	8.00%	36.54%	2.46%	0.00%	28.35%	0.00%	0.00%
15	0.33%	0.00%	2.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.85%	0.00%	2.07%	0.00%	0.00%
16	19.23%	0.00%	29.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	46.88%	46.60%	0.00%	57.68%	0.00%	0.00%
17	9.84%	0.00%	2.00%	0.00%	0.00%	0.00%	0.00%	16.95%	0.00%	0.00%	59.89%	0.00%	108.54%	0.00%	0.00%
18	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.13%	1.24%	0.00%	2.53%	0.00%	0.00%
19	0.25%	0.00%	10.96%	1.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.65%	0.00%	20.08%	0.00%	0.00%
20	0.00%	0.00%	0.00%	1.18%	0.00%	1.25%	0.00%	5.45%	0.00%	0.00%	10.66%	0.00%	18.74%	0.00%	0.00%
21	0.00%	0.00%	1.39%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.26%	0.00%	1.52%	0.00%	0.00%
22	0.55%	0.00%	16.42%	3.08%	0.00%	1.15%	0.00%	18.52%	2.17%	0.00%	27.04%	0.00%	69.49%	0.00%	0.00%
23	6.90%	0.00%	26.80%	1.45%	0.00%	8.14%	0.00%	14.29%	19.15%	1.96%	21.69%	0.00%	30.22%	0.00%	0.00%
24	5.81%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.82%	0.00%	2.22%	34.43%	0.00%	39.05%	0.00%	0.00%
25	47.64%	0.00%	5.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.42%	0.00%	2.93%	0.00%	0.00%
26	34.35%	0.00%	29.00%	0.00%	0.00%	0.00%	0.00%	3.53%	0.00%	40.00%	50.68%	0.00%	59.82%	0.00%	0.00%
27	1.87%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.53%	0.00%	0.00%	2.96%	0.00%	16.17%	0.00%	0.00%
28	0.31%	0.00%	9.09%	2.17%	0.00%	0.00%	0.00%	3.85%	0.00%	0.00%	15.45%	0.00%	23.26%	0.00%	0.00%
29	0.66%	0.00%	2.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	14.50%	0.00%	13.79%	0.00%	0.00%
30	0.00%	0.00%	2.06%	1.20%	0.00%	7.69%	0.00%	9.68%	4.17%	0.00%	1.23%	0.00%	13.18%	0.00%	0.00%
31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.43%	0.00%	0.62%	0.00%	0.00%
32	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.63%	0.00%	1.81%	0.00%	0.00%
33	1.11%	0.00%	8.05%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%	0.00%	3.95%	0.00%	6.79%	0.00%	0.00%
34	3.02%	0.00%	10.96%	0.00%	0.00%	1.06%	0.00%	4.65%	0.00%	0.00%	22.78%	0.00%	30.11%	0.00%	0.00%
35	0.00%	0.00%	1.41%	1.54%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%	10.00%	0.00%	12.78%	0.00%	0.00%
36	0.00%	0.00%	7.58%	1.30%	0.00%	0.00%	0.00%	8.51%	0.00%	0.00%	1.06%	0.00%	13.11%	0.00%	0.00%
37	0.00%	0.00%	0.00%	0.00%	0.00%	6.85%	0.00%	5.49%	9.09%	9.09%	26.47%	0.00%	34.13%	0.00%	0.00%
38	36.68%	0.00%	3.85%	1.37%	0.00%	4.69%	0.00%	15.79%	2.17%	8.16%	22.72%	0.00%	39.62%	0.00%	0.07%
39	0.40%	0.00%	4.71%	3.85%	0.00%	0.00%	0.00%	14.29%	0.00%	0.00%	28.98%	0.00%	57.87%	0.00%	0.00%
40	0.40%	0.00%	4.71%	3.85%	0.00%	0.00%	0.00%	14.29%	0.00%	0.00%	28.98%	0.00%	57.87%	0.00%	0.00%
41	1.72%	0.00%	7.87%	0.00%	0.00%	1.39%	0.00%	0.00%	0.00%	0.00%	7.98%	0.00%	9.08%	0.00%	0.00%
42	0.00%	0.00%	15.38%	0.00%	0.00%	0.00%	0.00%	10.53%	0.00%	0.00%	16.02%	0.00%	37.42%	0.00%	0.00%
43	5.15%	0.00%	15.38%	0.00%	0.00%	1.89%	0.00%	22.50%	0.00%	8.57%	19.49%	0.00%	39.80%	0.00%	0.00%
44	2.68%	0.00%	1.01%	0.00%	0.00%	1.00%	0.00%	0.00%	0.00%	0.00%	33.86%	0.00%	35.57%	0.00%	0.00%
45	2.56%	0.00%	9.09%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.93%	0.00%	11.65%	0.00%	0.00%
46	3.64%	0.00%	28.00%	0.00%	0.00%	7.14%	0.00%	10.71%	14.58%	44.00%	17.93%	0.00%	18.54%	0.00%	0.00%
47	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.26%	0.00%	0.00%
48	5.92%	0.00%	8.16%	0.00%	0.00%	6.90%	0.00%	0.00%	5.13%	12.12%	42.42%	0.00%	40.58%	0.00%	0.00%
49	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.24%	0.00%	3.61%	0.00%	0.00%
50	0.48%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.78%	1.89%	0.00%	2.64%	0.00%	0.00%
Arithmetic Mean	13.02%		11.73%	75.06%		1.65%		8.64%	1.71%	6.40%	21.37%		29.45%		
Median	1.03%		7.63%	0.00%		0.00%		3.69%	0.00%	0.00%	15.55%		19.26%		

Table 3.34: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\mathcal{P}_1, \ldots, \mathcal{P}_{50}\}$ generated from the LNP model while only observing A and I data. Also shown are the mean and median errors for each parameter. In this case, we unfix s_I from Table 3.24 which results in μ_{VT}, μ_{TI}, γ_T, β_{IB}, A_0, and I_0 being fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%.

By factoring out the s_I from the denominator, the term becomes

$$\frac{\beta_{BI}}{s_I}\left(\frac{I}{\frac{I}{s_I}+1}\right)B \tag{3.35}$$

If I/s_I is small, then the denominator becomes approximately 1 in which case

$$\frac{\beta_{BI}}{s_I} IB \tag{3.36}$$

Similar to what we did in the previous section with μ_{TB}/μ_{TI}, we can compare the exact values of the ratio β_{BI}/s_I to the fitted values of the ratio. If the errors of the ratio are consistently low even if the individual errors of β_{BI} and s_I are high, then it suggests that the parameters are, from a practical perspective, locally and not globally identifiable. We consider the five virtual patients with the highest errors in s_I which gives the sample $\{\mathcal{P}_2, \mathcal{P}_4, \mathcal{P}_{11}, \mathcal{P}_{17}, \mathcal{P}_{22}\}$. The comparison of the ratio values can be seen in Table 3.37.

Patient	β_{BI} exact	s_I exact	β_{BI}/s_I exact	β_{BI} fitted	s_I fitted	β_{BI}/s_I fitted	β_{BI} error	s_I error	β_{BI}/s_I error
$\mathcal{P}_2$	4.61	1000	0.00461	7.49	1658.24	0.00452	62.47%	65.82%	1.95%
$\mathcal{P}_4$	2.62	1000	0.00262	4.71	2736.49	0.00172	79.77%	173.65%	34.35%
$\mathcal{P}_{11}$	2.63	1000	0.00263	4.06	1829.18	0.00222	54.37%	82.92%	15.59%
$\mathcal{P}_{17}$	3.59	1000	0.00359	5.74	2085.42	0.00275	59.89%	108.54%	23.40%
$\mathcal{P}_{22}$	2.33	1000	0.00233	2.96	1694.92	0.00175	27.04%	69.49%	24.89%

Table 3.37: Comparison of the exact MATLAB values, Monolix fitted values, and errors for β_{BI}, s_I, and the ratio β_{BI}/s_I given the virtual patient sample $\{\mathcal{P}_2, \mathcal{P}_4, \mathcal{P}_{11}, \mathcal{P}_{17}, \mathcal{P}_{22}\}$. Note that the ratio vales are rounded to 3 significant figures while the ratio errors are rounded to 2 decimal places.

We see that for all five patients, the ratio errors of β_{BI}/s_I are lower than those for the individual errors of β_{BI} and s_I. Referring back to equation 3.35, we note that when $I \ll s_I$, the I/s_I term approaches zero allowing us to write the asymptotic approximation

$$\frac{\beta_{BI}}{s_I}\left(\frac{I}{\frac{I}{s_I}+1}\right)B \sim \frac{\beta_{BI}}{s_I} IB, \quad I \ll s_I \tag{3.38}$$

This effectively removes the denominator of the non–linear ratio term leaving us with β_{BI}/s_I as we previously showed. To confirm whether or not $I \ll s_I$, we can consider for these five patients the mean values of I across the logarithmically spaced data points divided by $s_I = 1000$. These values are displayed in Table 3.39. We see that the ratios between the average values of I and s_I are roughly $\mathcal{O}(10^{-1})$. This results in the ratio term being approximately $I/1.1$ and so in this case, we can more confidently say that the asymptotic approximation in equation 3.38 is valid.

Patient	Mean$(I)/s_I$
$\mathcal{P}_2$	0.102
$\mathcal{P}_4$	0.242
$\mathcal{P}_{11}$	0.235
$\mathcal{P}_{17}$	0.170
$\mathcal{P}_{22}$	0.293

Table 3.39: Mean values of I divided by $s_I = 1000$ from the datasets $\mathcal{U}_i$ of the virtual patient sample $\{\mathcal{P}_2, \mathcal{P}_4, \mathcal{P}_{11}, \mathcal{P}_{17}, \mathcal{P}_{22}\}$. Note that the values are rounded to 3 significant figures.

If s_I was fixed like before, then β_{BI} would be globally identifiable. However since it is variable, both parameters become locally identifiable as there are infinitely many solutions to the ratio. We therefore suspect that similar to the case of μ_{TB}/μ_{TI} from Section 3.3, the poorly fitted values for β_{BI} and s_I are the result of the ratio being better identified than the individual parameter values. Furthermore, we have $\beta_{BI}/s_I = \varepsilon \sim \mathcal{O}(10^{-3})$. If s_I is fixed, then we identify β_{BI} from $\varepsilon s_I \sim \mathcal{O}(1)$ whereas if β_{BI} is fixed, we identify s_I from $\beta_{BI}/\varepsilon \sim \mathcal{O}(10^3)$. Once again, the terms are more sensibly scaled if s_I is fixed rather than β_{BI}. Given this, we maintain the fixing of s_I for the remainder of this chapter.

3.5 Identifiability analysis of a population

So far, we have only fitted patients individually by providing single patient datasets in Sections 2.4, 2.5, and 2.6 and by providing a population of five identical patient datasets from Section 2.7 onwards. Of course, one of the main intended uses of the SAEM algorithm in Monolix is to analyse a varied patient population where all datasets are provided at the same time. In doing so, Monolix outputs parameter fits both of the patient population as a whole (these are equivalent to the "individual" parameter fits from the previous sections) as well as individual fits for each patient separately. Naturally, we are interested in employing Monolix in this way and so we close out this chapter by considering cases where the entire patient population is fitted together. Since we are now inputting different patients instead of identical patients for a population fit, we do not need to concern ourselves with the identical patient populations in $\{\mathcal{P}_1, ..., \mathcal{P}_{50}\}$. Instead, we can go back to the patient population $\{P_1, ..., P_{50}\}$ from Section 2.6 where data points were not repeated and simply input all patients into Monolix at the same time. When doing this, our data reflects the case from Section 3.2 where we set only A and I as observed state variables. The main disadvantage of this is the fact that parameter fixing in Monolix is done at the population level and not the individual level. Hence, we cannot fix individual patient parameters since they are all

varied across the population. Therefore, with the exception of s_I which is the same value for all patients, all parameters are kept unfixed. So, now considering $\{P_1, \ldots, P_{50}\}$ as a single inputted population P where $P_i = (\mathbf{p}_i, U_i)$, the experimental procedure is as follows

1. Load the solution dataset U attained by concatenating together all of the U_i's
2. Set A, and I as observed state variables
3. Load the LNP model equations
4. Fix $s_I = 1000$ with random effects disabled
5. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution.
6. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
7. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
8. Run the SAEM algorithm which outputs the population parameter fits as well as the individual parameter fits which in this case are what we use to compare the parameter values for each patient

The results of the individual parameter fits are shown in Table 3.40. As we can see, referring back to the DAISY identifiability classifications from Table 3.17, the globally identifiable decay rates γ_B, γ_A, and γ_I are fitted better than the locally identifiable γ_V and γ_T. Furthermore, barring μ_{LV} which is globally identifiable but is also as far upstream as possible from the observed state variables, all of the non–identifiable priming rates μ_{ij} are fitted poorly. Similarly, the globally identifiable autocatalytic rate β_{BI} is fitted well while the non–identifiable β_{IB} is also fitted poorly. Finally, unsurprisingly the initial value A_0 and I_0 are fitted almost exactly for all patients since the individual datasets all have data points at $t = 0$. Of course, the overall mean error is 120.62% which is high compared to the case where we fixed parameters in Table 3.24 and helps to justify our use of the identical patient populations for fitting individual patients in the previous sections as it allowed us to fix parameters.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	52.74%	33.88%	200.00%	37.04%	123.26%	16.67%	34.02%	13.56%	0.00%	0.00%	15.84%	65.63%	0.00%	0.25%	0.23%
2	98.27%	374.95%	38.46%	100.00%	435.80%	1917.24%	16.67%	50.00%	48.98%	23.08%	5.64%	185.71%	0.00%	1.42%	0.37%
3	98.10%	932.47%	16.18%	178.43%	151.37%	976.92%	120.93%	5.66%	0.00%	4.35%	4.36%	252.94%	0.00%	0.00%	0.96%
4	12.56%	38.76%	7.14%	141.67%	79.85%	7.00%	16.67%	5.88%	0.00%	0.00%	4.58%	141.18%	0.00%	0.16%	0.75%
5	96.33%	122.49%	1050.00%	5.56%	878.12%	983.33%	85.11%	0.00%	2.56%	18.92%	0.27%	0.00%	0.00%	0.01%	0.91%
6	30.34%	35.86%	56.15%	326.79%	101.75%	20.00%	4.17%	4.26%	0.00%	22.22%	3.41%	376.92%	0.00%	0.05%	0.00%
7	0.00%	17.85%	52.00%	156.25%	22.75%	4.29%	2.99%	0.00%	0.00%	0.00%	0.74%	155.56%	0.00%	0.08%	0.91%
8	10.00%	13.94%	49.17%	143.24%	9.17%	47.27%	78.57%	0.00%	0.00%	1.85%	0.50%	140.91%	0.00%	0.10%	1.02%
9	28.57%	16.08%	64.55%	229.41%	10.47%	6.90%	27.27%	0.00%	0.00%	0.00%	1.53%	263.64%	0.00%	0.09%	0.72%
10	38.46%	441.07%	81.58%	895.96%	57.64%	9.09%	54.55%	30.61%	0.00%	2.56%	38.46%	1030.43%	0.00%	0.46%	1.00%
11	48.15%	286.91%	78.57%	140.00%	47.66%	92.86%	1.79%	0.00%	0.00%	0.00%	0.76%	145.16%	0.00%	0.10%	1.07%
12	9.16%	60.65%	53.85%	48.65%	31.68%	12.50%	7.79%	1.85%	0.00%	2.56%	1.96%	50.00%	0.00%	0.11%	0.73%
13	75.71%	845.00%	28.46%	58.06%	72.87%	118.18%	8.33%	38.00%	0.00%	0.00%	45.95%	37.50%	0.00%	0.13%	0.65%
14	90.55%	246.87%	76.92%	56.67%	178.94%	696.30%	20.00%	0.00%	0.00%	5.77%	0.82%	69.23%	0.00%	0.08%	0.66%
15	87.04%	37.56%	48.00%	43.33%	104.71%	83.33%	69.17%	85.71%	110.53%	96.25%	61.54%	18.42%	0.00%	3.46%	0.61%
16	35.90%	287.95%	65.22%	22.22%	75.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.81%	0.00%	0.07%	1.08%
17	98.22%	919.33%	56.00%	476.92%	135.47%	1900.00%	7.59%	3.39%	0.00%	0.00%	2.23%	553.85%	0.00%	0.04%	0.35%
18	60.18%	6.13%	225.00%	6.19%	39.37%	68.00%	92.31%	36.00%	316.67%	176.60%	20.43%	13.64%	0.00%	3.34%	3.13%
19	21.43%	192.86%	75.34%	97.92%	60.02%	34.62%	70.45%	5.56%	0.00%	0.00%	4.76%	96.97%	0.00%	0.16%	0.54%
20	10.54%	241.79%	64.29%	10.59%	69.58%	12.50%	7.79%	1.82%	0.00%	0.00%	0.21%	9.68%	0.00%	0.11%	0.70%
21	84.72%	262.81%	13.89%	100.00%	64.39%	480.00%	5.00%	0.00%	0.00%	0.00%	1.52%	111.76%	0.00%	0.10%	0.72%
22	1.94%	5.40%	67.16%	189.23%	5.78%	50.57%	104.65%	0.00%	0.00%	2.17%	0.43%	187.88%	0.00%	0.08%	0.91%
23	91.95%	81.47%	157.73%	128.99%	412.69%	2.33%	750.00%	1.43%	0.00%	1.96%	1.84%	160.87%	0.00%	0.12%	0.16%
24	85.81%	531.64%	33.33%	59.57%	53.33%	188.46%	33.87%	3.64%	0.00%	35.56%	4.64%	39.29%	0.00%	0.09%	0.65%
25	94.09%	1154.42%	71.62%	139.13%	53.21%	481.63%	8.97%	3.23%	0.00%	0.00%	3.54%	174.07%	0.00%	0.10%	0.34%
26	80.61%	9.37%	28.00%	358.82%	112.33%	92.31%	2.04%	7.06%	0.00%	7.50%	8.17%	450.00%	0.00%	0.13%	0.14%
27	92.16%	148.63%	7.69%	226.47%	134.56%	38.33%	683.13%	5.26%	0.00%	2.86%	2.96%	307.14%	0.00%	0.08%	0.49%
28	95.28%	614.86%	18.18%	35.87%	74.71%	65.38%	3077.78%	3.85%	0.00%	0.00%	1.82%	60.00%	0.00%	0.07%	0.47%
29	23.18%	319.66%	91.72%	320.37%	73.70%	27.27%	13.64%	5.36%	0.00%	0.00%	6.00%	317.39%	0.00%	0.17%	0.59%
30	88.89%	9.43%	889.69%	27.71%	506.33%	561.54%	2.50%	3.23%	2.08%	0.00%	1.72%	25.00%	0.00%	0.00%	0.36%
31	88.74%	74.62%	135.71%	54.64%	223.58%	350.00%	24.10%	20.00%	0.00%	25.81%	25.80%	21.74%	0.00%	0.66%	3.63%
32	8.46%	21.47%	14.29%	53.73%	32.98%	0.00%	0.00%	0.00%	0.00%	0.00%	1.87%	60.53%	0.00%	0.03%	0.71%
33	74.52%	505.15%	26.44%	18.75%	64.84%	300.00%	31.18%	28.13%	0.00%	10.81%	26.58%	33.33%	0.00%	0.43%	0.76%
34	97.39%	4043.87%	9.59%	5.56%	56.35%	17.02%	1176.60%	9.30%	0.00%	3.13%	10.31%	35.48%	0.00%	0.15%	0.05%
35	95.87%	457.43%	1660.56%	84.62%	146.32%	60.83%	2878.72%	0.00%	0.00%	3.23%	1.06%	81.84%	0.00%	0.00%	0.10%
36	68.42%	172.20%	71.21%	327.27%	18.70%	227.27%	2.38%	0.00%	0.00%	4.08%	0.64%	324.24%	0.00%	0.09%	0.86%
37	96.77%	606.25%	29.29%	326.92%	224.36%	1995.89%	1.85%	2.20%	0.00%	3.03%	2.52%	354.55%	0.00%	0.02%	0.22%
38	98.77%	538.04%	361.54%	23.29%	260.67%	56.25%	1014.75%	2.63%	0.00%	26.53%	3.34%	8.33%	0.00%	0.02%	0.83%
39	63.64%	284.16%	36.92%	7.69%	35.45%	47.62%	219.15%	5.71%	0.00%	15.38%	3.49%	60.71%	0.00%	0.10%	0.43%
40	29.64%	7.94%	31.76%	151.92%	44.41%	57.86%	183.33%	7.14%	0.00%	2.22%	6.37%	170.27%	0.00%	0.13%	0.47%
41	87.41%	275.60%	41.57%	295.65%	113.46%	719.44%	3.85%	4.65%	4.76%	11.43%	0.25%	305.41%	0.00%	0.05%	0.84%
42	82.86%	288.82%	24.62%	11.11%	26.34%	6.41%	342.11%	3.51%	0.00%	3.57%	4.33%	6.45%	0.00%	0.06%	0.47%
43	23.71%	196.06%	62.31%	11.11%	60.92%	1.89%	9.09%	0.00%	0.00%	2.86%	0.51%	16.67%	0.00%	0.07%	0.69%
44	93.36%	289.58%	15.15%	210.00%	128.90%	640.00%	16.88%	2.94%	0.00%	0.00%	3.59%	214.29%	0.00%	0.09%	0.60%
45	38.97%	899.29%	88.18%	65.31%	85.90%	49.29%	163.16%	6.67%	0.00%	3.13%	5.91%	75.00%	0.00%	0.13%	0.51%
46	3.64%	18.23%	180.00%	68.75%	11.52%	37.14%	25.53%	0.00%	0.00%	12.00%	1.09%	67.74%	0.00%	0.13%	0.69%
47	31.15%	50.09%	26.67%	232.81%	140.80%	25.83%	53.85%	0.00%	0.00%	0.00%	3.08%	253.85%	0.00%	0.06%	0.36%
48	98.28%	982.37%	30.61%	345.45%	225.45%	3055.17%	0.00%	0.00%	0.00%	3.03%	0.43%	400.00%	0.00%	0.07%	0.04%
49	0.00%	27.11%	76.43%	523.38%	37.33%	16.67%	8.33%	0.00%	0.00%	0.00%	0.50%	511.11%	0.00%	0.07%	0.89%
50	91.79%	3090.87%	84.55%	4.88%	44.51%	484.42%	70.83%	33.33%	233.33%	81.94%	15.77%	341.18%	0.00%	5.62%	3.15%
Arithmetic Mean	60.29%	422.39%	135.47%	151.68%	123.19%	342.88%	232.67%	8.83%	14.38%	12.33%	7.28%	176.19%		0.38%	0.75%
Median	75.11%	244.33%	56.08%	98.96%	73.29%	57.05%	24.81%	3.31%	0.00%	2.71%	3.02%	126.34%		0.10%	0.66%

Table 3.40: Percentage errors of individual parameter fits attained from simultaneously fitting the full patient population $\{P_1, \ldots, P_{50}\}$ generated from the LNP model while only observing A and I data. Also shown are the mean and median errors for each parameter. In this case $s_I = 1000$ is fixed to its exact value for the whole population. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%.

We therefore consider a way to fit an entire virtual patient population simultaneously while also being able to fix parameters informed by all of the analysis from Chapters 2 and 3 thus far. In the previous experiment, we were able to fix s_I as it was set to the same value for all virtual patients $P_1, \ldots, P_{50}$ when they were generated. So we can generate a new patient population where param-

eters that we wish to fix are set to the same value for all patients. We take the fixed parameter set from Table 3.24 barring the initial conditions A_0 and I_0 since those can effectively be kept variable given the previous results. Therefore, the fixed parameter set is $\{\mu_{VT},\ \mu_{TI},\ \gamma_V,\ \beta_{IB},\ s_I\}$. For all patients, we fix each of these parameters to the expected values (midpoints) of their respective distributions from Table 2.5. This gives us the following fixed values

$$\mu_{VT} = 15.1, \quad \mu_{TI} = 14.19, \quad \gamma_T = 0.08, \quad \beta_{IB} = 0.0025, \quad s_I = 1000$$

We maintain the time point vector $\mathbf{t}$ consisting of 101 points and generate a new population of 50 virtual patients $\widehat{P}_i = (\widehat{\mathbf{p}}_i,\ \widehat{U}_i)$ under the following scheme

1. Generate 50 parameter vectors from the LNP model in equations 2.20a – 2.20f uniformly using MATLAB's `rand` function with the distributions seen in Table 2.5 which yields the parameter vector $\widehat{\mathbf{p}}_i$

   (a) The following parameters are fixed to the same value for all patients
   $\mu_{VT} = 15.1, \quad \mu_{TI} = 14.19, \quad \gamma_T = 0.08, \quad \beta_{IB} = 0.0025, \quad s_I = 1000$

2. For each $\widehat{\mathbf{p}}_i$ generate the solution matrix $\widehat{U}_i'$ using the `ode45` solver given the model equations

3. The matrices $\widehat{U}_i'$ use the adaptive time stepping scheme and so regenerate them using `spline` given the timespan vector $\mathbf{t}$ consisting of one time point at $t = 0$ and 100 logarithmically spaced time points from $t = 0.001$ to $t = 365$ which gives $\widehat{U}_i = \begin{pmatrix} \mathbf{L} & \mathbf{V} & \mathbf{T} & \mathbf{B} & \mathbf{A} & \mathbf{I} \end{pmatrix}$

4. Group the parameter vectors and splined solution matrices together as $\widehat{P}_i = \{\widehat{\mathbf{p}}_i,\ \widehat{U}_i\}$ yielding the virtual patients $\widehat{P}_1,\ ...,\ \widehat{P}_{50}$

The individual parameter values for the patient population $\{\widehat{P}_1,\ ...,\ \widehat{P}_{50}\}$ are shown in Table 3.41.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	1.64	15.1	0.13	0.21	14.19	0.071	0.08	0.49	0.038	0.046	3.96	0.0025	1000	566.01	1.49
2	0.38	15.1	0.079	0.82	14.19	0.054	0.08	0.51	0.037	0.037	3.03	0.0025	1000	433.13	121.64
3	1.13	15.1	0.07	0.32	14.19	0.11	0.08	0.46	0.038	0.029	3.56	0.0025	1000	311.72	189.17
4	2.22	15.1	0.14	0.99	14.19	0.13	0.08	0.46	0.049	0.049	2.99	0.0025	1000	190.6	106.37
5	3.3	15.1	0.12	0.36	14.19	0.1	0.08	0.32	0.035	0.031	3.04	0.0025	1000	285.51	73.13
6	0.2	15.1	0.065	0.69	14.19	0.075	0.08	0.5	0.044	0.022	3.06	0.0025	1000	398.95	253.99
7	0.94	15.1	0.11	1.02	14.19	0.11	0.08	0.31	0.049	0.029	4.6	0.0025	1000	484.31	257.06
8	2.72	15.1	0.072	0.99	14.19	0.12	0.08	0.17	0.052	0.054	4.36	0.0025	1000	174.08	104.76
9	0.87	15.1	0.1	0.53	14.19	0.067	0.08	0.35	0.04	0.021	3.71	0.0025	1000	60.8	58.34
10	1.93	15.1	0.095	0.54	14.19	0.13	0.08	0.56	0.041	0.031	4.88	0.0025	1000	94.93	22.31
11	1.02	15.1	0.14	0.081	14.19	0.13	0.08	0.046	0.049	0.024	3.49	0.0025	1000	390.21	126.48
12	0.37	15.1	0.12	0.75	14.19	0.12	0.08	0.28	0.035	0.021	4.64	0.0025	1000	435.62	28.5
13	3.12	15.1	0.09	0.36	14.19	0.1	0.08	0.41	0.047	0.047	3.53	0.0025	1000	168.74	77.01
14	1.03	15.1	0.097	0.76	14.19	0.052	0.08	0.18	0.051	0.047	2.83	0.0025	1000	537.05	38.36
15	1.68	15.1	0.13	0.1	14.19	0.13	0.08	0.26	0.044	0.027	4.15	0.0025	1000	322.75	125.86
16	1.14	15.1	0.088	0.79	14.19	0.061	0.08	0.069	0.036	0.039	4.16	0.0025	1000	549.87	198.9
17	1.06	15.1	0.099	0.33	14.19	0.077	0.08	0.11	0.04	0.042	3.13	0.0025	1000	99.86	120.52
18	3.12	15.1	0.1	0.17	14.19	0.069	0.08	0.43	0.036	0.023	4.68	0.0025	1000	111.2	192.69
19	3.24	15.1	0.14	0.91	14.19	0.087	0.08	0.31	0.039	0.02	2.51	0.0025	1000	352.16	199.75
20	3.35	15.1	0.079	0.24	14.19	0.091	0.08	0.52	0.042	0.055	3.09	0.0025	1000	232.26	120.12
21	3.09	15.1	0.14	0.41	14.19	0.11	0.08	0.027	0.038	0.026	3.7	0.0025	1000	457.98	91.46
22	3.48	15.1	0.067	0.5	14.19	0.063	0.08	0.16	0.035	0.054	4.46	0.0025	1000	136.97	163.77
23	3.06	15.1	0.081	0.48	14.19	0.13	0.08	0.065	0.04	0.048	3.19	0.0025	1000	386.75	67.67
24	2.76	15.1	0.14	0.95	14.19	0.13	0.08	0.37	0.038	0.048	4.17	0.0025	1000	136.85	52.75
25	2.79	15.1	0.065	0.4	14.19	0.054	0.08	0.29	0.041	0.023	4.49	0.0025	1000	399.61	197.18
26	2.72	15.1	0.076	0.18	14.19	0.076	0.08	0.081	0.042	0.042	2.96	0.0025	1000	37.18	231.72
27	0.68	15.1	0.084	0.66	14.19	0.064	0.08	0.48	0.042	0.029	4.3	0.0025	1000	491.94	90.86
28	1.96	15.1	0.069	1.01	14.19	0.08	0.08	0.46	0.046	0.033	3.42	0.0025	1000	119.38	207.17
29	4	15.1	0.067	0.4	14.19	0.13	0.08	0.5	0.049	0.032	3.97	0.0025	1000	81.63	26.18
30	3.78	15.1	0.084	0.53	14.19	0.1	0.08	0.083	0.049	0.026	3.92	0.0025	1000	144.69	110.6
31	2.69	15.1	0.13	0.75	14.19	0.1	0.08	0.13	0.052	0.024	2.26	0.0025	1000	447.07	89.73
32	4.01	15.1	0.13	0.46	14.19	0.054	0.08	0.34	0.052	0.02	3.15	0.0025	1000	338.87	180.6
33	2.79	15.1	0.13	0.85	14.19	0.057	0.08	0.44	0.046	0.052	2	0.0025	1000	46.96	230.14
34	2.51	15.1	0.12	0.51	14.19	0.081	0.08	0.1	0.052	0.046	3.09	0.0025	1000	10.11	127.7
35	1.19	15.1	0.11	0.63	14.19	0.074	0.08	0.18	0.052	0.042	4.79	0.0025	1000	81.61	111.74
36	3.89	15.1	0.089	0.26	14.19	0.095	0.08	0.35	0.044	0.039	4.2	0.0025	1000	170.08	42.01
37	2.56	15.1	0.094	0.97	14.19	0.11	0.08	0.41	0.045	0.036	1.95	0.0025	1000	268.29	226.32
38	0.79	15.1	0.096	0.67	14.19	0.092	0.08	0.042	0.037	0.029	2.56	0.0025	1000	164.17	65.37
39	2.35	15.1	0.081	0.61	14.19	0.11	0.08	0.18	0.052	0.023	4.05	0.0025	1000	171.01	118.77
40	1.88	15.1	0.11	0.25	14.19	0.093	0.08	0.18	0.036	0.025	3.57	0.0025	1000	48.79	129.58
41	1.84	15.1	0.098	0.28	14.19	0.083	0.08	0.26	0.038	0.03	3.95	0.0025	1000	331.52	205.67
42	1.63	15.1	0.13	0.43	14.19	0.084	0.08	0.056	0.047	0.043	3.6	0.0025	1000	322.69	177.07
43	0.78	15.1	0.079	0.52	14.19	0.083	0.08	0.41	0.044	0.026	2.51	0.0025	1000	307.43	200.46
44	3.92	15.1	0.083	0.26	14.19	0.11	0.08	0.21	0.041	0.037	2.51	0.0025	1000	3.04	79.2
45	3.58	15.1	0.11	0.66	14.19	0.11	0.08	0.44	0.041	0.027	2.83	0.0025	1000	144.84	31.48
46	3.42	15.1	0.085	0.13	14.19	0.14	0.08	0.49	0.051	0.046	3.21	0.0025	1000	49.97	95.75
47	2.82	15.1	0.12	0.59	14.19	0.096	0.08	0.24	0.049	0.028	4.61	0.0025	1000	263.47	139.44
48	3.94	15.1	0.093	0.2	14.19	0.071	0.08	0.59	0.05	0.034	3.97	0.0025	1000	266.37	112.79
49	2.3	15.1	0.066	0.39	14.19	0.12	0.08	0.066	0.046	0.027	2.75	0.0025	1000	374.66	220.81
50	2.26	15.1	0.13	0.088	14.19	0.055	0.08	0.14	0.045	0.03	4.15	0.0025	1000	476.28	131.62

Table 3.41: Parameter values of virtual patients $\widehat{P}_1$, ..., $\widehat{P}_{50}$ generated in MATLAB for the LNP model. Note that $\mu_{VT} = 15.1$, $\mu_{TI} = 14.19$, $\gamma_T = 0.08$, $\beta_{IB} = 0.0025$, and $s_I = 1000$ are fixed for all patients.

We now proceed with the experiment in Monolix which is essentially identical to the previous one but with the new patient population and fixed parameters. The experimental procedure is as follows

1. Load the solution dataset $\widehat{U}$ attained by concatenating together all of the $\widehat{U}_i$'s

2. Set A, and I as observed state variables

3. Load the LNP model equations

4. Fix $\mu_{VT} = 15.1,\ \mu_{TI} = 14.19,\ \gamma_T = 0.08,\ \beta_{IB} = 0.0025,$ and $s_I = 1000$ with random effects disabled

5. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution.

6. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space

7. Set posterior distributions for data on all compartments to normal distributions (Monolix default)

8. Run the SAEM algorithm which outputs the population parameter fits as well as the individual parameter fits which in this case are what we use to compare the parameter values for each patient

The results can be seen in Table 3.42. The mean error is 7.22% which is a significant improvement over the 120.62% mean error from Table 3.40 and almost as low as the individual patient fits mean error from Table 3.23 which was 2.47%. This once again emphasises the conclusions of Sections 3.1 and 3.2; namely that sufficient T–cell and interleukin data is very useful for identifiability if it allows for a fixing of parameters even on a population level. Notably, instead of the higher errors being more consistently distributed across certain parameters, they are instead concentrated on a select few patients. This suggests that there may be a propagation where the SAEM algorithm is correcting the fits of other patients by incorrectly identifying others. We can verify this by eliminating the badly fit patients which in this case we consider to be $\widehat{P}_{18}$, $\widehat{P}_{21}$, $\widehat{P}_{22}$, $\widehat{P}_{23}$, $\widehat{P}_{32}$, and $\widehat{P}_{46}$. The experiment is the exact same with the population simply modified to have said patients removed and the results of the parameter fits can be seen in Table 3.43. We see that the parameter fits actually seem to be worse and in fact, the mean error is higher at 13.99%. Perhaps most importantly, $\widehat{P}_{25}$, $\widehat{P}_{26}$, and $\widehat{P}_{27}$ are now some of the worst fitted patients despite being very well fitted in Table 3.42. This confirms two points. Firstly, under such patient populations where the parameters with low identifiability are set to the same value such that they can be fixed, at least some error propagation is present and patients that are not fitted well may improve if fitted separately. Secondly, a larger population size seems to reduce the effects of the error propagation resulting in a lower mean

error. Therefore, we can conclude that the effectiveness of our identifiability analysis in Monolix can be improved by maximising the population size.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.01%	2.68%
2	0.00%	0.00%	2.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.33%	0.00%	0.00%	0.01%	0.02%
3	1.77%	0.00%	4.29%	0.00%	0.00%	0.00%	0.00%	2.17%	0.00%	0.00%	1.40%	0.00%	0.00%	0.02%	0.01%
4	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.03%
5	0.61%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.66%	0.00%	0.00%	0.01%	0.11%
6	0.00%	0.00%	1.54%	1.45%	0.00%	1.33%	0.00%	0.00%	0.00%	0.00%	0.33%	0.00%	0.00%	0.01%	0.07%
7	2.13%	0.00%	0.00%	0.98%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.03%	0.00%
8	1.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.23%	0.00%	0.00%	0.02%	0.03%
9	1.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.10%
10	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	0.00%	0.00%	0.01%	0.09%
11	4.90%	0.00%	7.14%	1.23%	0.00%	0.00%	0.00%	2.17%	2.04%	50.00%	1.72%	0.00%	0.00%	0.07%	0.06%
12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.65%	0.00%	0.00%	0.00%	0.04%
13	1.60%	0.00%	1.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.57%	0.00%	0.00%	0.02%	0.03%
14	0.00%	0.00%	1.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.13%	0.00%	0.00%	0.00%	0.01%	0.03%
15	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.05%
16	1.75%	0.00%	3.41%	0.00%	0.00%	1.64%	0.00%	1.45%	0.00%	12.82%	0.48%	0.00%	0.00%	0.01%	0.20%
17	0.94%	0.00%	2.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.76%	0.64%	0.00%	0.00%	0.00%	0.06%
18	27.88%	0.00%	30.00%	0.00%	0.00%	56.52%	0.00%	13.95%	166.67%	465.22%	1.71%	0.00%	0.00%	2.70%	1.66%
19	0.31%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	0.02%	0.10%
20	1.49%	0.00%	1.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.02%
21	3.88%	0.00%	21.43%	4.88%	0.00%	0.00%	0.00%	40.74%	28.95%	42.31%	5.68%	0.00%	0.00%	0.48%	0.02%
22	33.05%	0.00%	79.10%	2.00%	0.00%	46.03%	0.00%	25.00%	142.86%	44.44%	6.05%	0.00%	0.00%	1.50%	0.24%
23	3.92%	0.00%	60.49%	18.75%	0.00%	15.38%	0.00%	35.38%	45.00%	35.42%	12.23%	0.00%	0.00%	0.77%	0.15%
24	0.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.02%	0.15%
25	0.00%	0.00%	3.08%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.45%	0.00%	0.00%	0.02%	0.04%
26	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.38%	0.00%	0.00%	0.00%	0.03%	0.05%
27	1.47%	0.00%	7.14%	0.00%	0.00%	0.00%	0.00%	2.08%	0.00%	0.00%	1.40%	0.00%	0.00%	0.03%	0.13%
28	0.00%	0.00%	1.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.29%	0.00%	0.00%	0.03%	0.07%
29	0.25%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.01%	0.08%
30	1.32%	0.00%	2.38%	0.00%	0.00%	1.00%	0.00%	1.20%	0.00%	7.69%	0.51%	0.00%	0.00%	0.01%	0.08%
31	0.37%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.17%	0.44%	0.00%	0.00%	0.01%	0.04%
32	42.39%	0.00%	53.85%	8.70%	0.00%	12.96%	0.00%	11.76%	7.69%	15.00%	14.29%	0.00%	0.00%	0.38%	0.97%
33	0.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.92%	0.00%	0.00%	0.00%	0.02%	0.05%
34	0.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.17%	0.00%	0.00%	0.00%	0.10%	0.12%
35	0.84%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.38%	0.21%	0.00%	0.00%	0.02%	0.05%
36	1.80%	0.00%	7.87%	0.00%	0.00%	1.05%	0.00%	2.86%	0.00%	2.56%	1.90%	0.00%	0.00%	0.05%	57.25%
37	0.78%	0.00%	1.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.51%	0.00%	0.00%	0.00%	0.04%
38	1.27%	0.00%	1.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	17.24%	0.00%	0.00%	0.00%	0.03%	0.00%
39	0.43%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.04%
40	0.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%
41	1.63%	0.00%	4.08%	0.00%	0.00%	1.20%	0.00%	0.00%	0.00%	3.33%	1.27%	0.00%	0.00%	0.03%	0.01%
42	0.00%	0.00%	7.69%	2.33%	0.00%	2.38%	0.00%	5.36%	4.26%	9.30%	1.67%	0.00%	0.00%	0.04%	0.15%
43	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%
44	0.26%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.64%	0.13%
45	0.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.35%	0.00%	0.00%	0.02%	0.00%
46	59.06%	0.00%	794.12%	46.15%	0.00%	62.14%	0.00%	59.18%	350.98%	117.39%	57.01%	0.00%	0.00%	3.48%	0.49%
47	1.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.22%	0.00%	0.00%	0.02%	0.01%
48	0.51%	0.00%	4.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%	0.00%	0.00%	0.02%	0.04%
49	4.35%	0.00%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	18.52%	0.36%	0.00%	0.00%	0.01%	0.05%
50	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.48%	0.00%	0.00%	0.02%	0.04%
Arithmetic Mean	4.18%		22.10%	1.73%		4.03%		4.07%	14.97%	17.29%	2.32%			0.24%	1.32%
Median	0.75%		1.04%	0.00%		0.00%		0.00%	0.00%	0.00%	0.36%			0.02%	0.06%

Table 3.42: Percentage errors of individual parameter fits attained from simultaneously fitting the full patient population $\{\widehat{P}_1, \ldots, \widehat{P}_{50}\}$ generated from the LNP model while only observing A and I data. Also shown are the mean and median errors for each parameter. In this case $\mu_{VT} = 15.1$, $\mu_{TI} = 14.19$, $\gamma_T = 0.08$, $\beta_{IB} = 0.0025$, and $s_I = 1000$ are fixed to its exact value for the whole population. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	0.61%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	2.04%	0.00%	0.00%	1.52%	0.00%	0.00%	0.08%	0.67%
2	26.32%	0.00%	51.90%	4.88%	0.00%	14.81%	0.00%	9.80%	8.11%	97.30%	10.56%	0.00%	0.00%	0.06%	0.71%
3	0.88%	0.00%	2.86%	3.13%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.28%	0.00%	0.00%	0.15%	0.00%
4	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.12%
5	14.55%	0.00%	275.00%	5.56%	0.00%	32.00%	0.00%	25.00%	151.43%	74.19%	21.71%	0.00%	0.00%	2.39%	0.27%
6	70.00%	0.00%	23.08%	13.04%	0.00%	26.67%	0.00%	14.00%	31.82%	300.00%	10.78%	0.00%	0.00%	1.66%	0.65%
7	1.06%	0.00%	9.09%	0.98%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.09%	0.00%	0.00%	0.05%	0.11%
8	1.10%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.69%	0.00%	0.00%	0.13%	0.14%
9	1.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.27%	0.00%	0.00%	0.05%	0.26%
10	0.00%	0.00%	2.11%	1.85%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.41%	0.00%	0.00%	0.20%	0.22%
11	5.88%	0.00%	7.14%	1.23%	0.00%	0.00%	0.00%	4.35%	4.08%	62.50%	1.72%	0.00%	0.00%	0.18%	0.14%
12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.18%
13	1.28%	0.00%	0.00%	0.00%	0.00%	1.00%	0.00%	0.00%	0.00%	2.13%	0.28%	0.00%	0.00%	0.03%	0.25%
14	0.97%	0.00%	2.06%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.71%	0.00%	0.00%	0.02%	0.23%
15	0.60%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.96%	0.00%	0.00%	0.08%	0.21%
16	1.75%	0.00%	1.14%	1.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.72%	0.00%	0.00%	0.11%	0.24%
17	0.94%	0.00%	1.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.64%	0.00%	0.00%	0.04%	0.14%
19	2.16%	0.00%	0.00%	1.10%	0.00%	1.15%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	0.10%	0.16%
20	0.30%	0.00%	5.06%	0.00%	0.00%	1.10%	0.00%	1.92%	0.00%	0.00%	0.97%	0.00%	0.00%	0.13%	0.05%
24	1.09%	0.00%	0.00%	2.11%	0.00%	0.00%	0.00%	0.00%	0.00%	2.08%	0.96%	0.00%	0.00%	0.07%	0.23%
25	73.48%	0.00%	207.69%	2.50%	0.00%	27.78%	0.00%	24.14%	58.54%	152.17%	19.15%	0.00%	0.00%	1.41%	0.26%
26	21.69%	0.00%	28.95%	5.56%	0.00%	18.42%	0.00%	18.52%	69.05%	45.24%	6.08%	0.00%	0.00%	1.05%	0.08%
27	19.12%	0.00%	911.90%	31.82%	0.00%	21.88%	0.00%	33.33%	85.71%	148.28%	36.28%	0.00%	0.00%	0.75%	0.04%
28	1.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.14%	0.31%
29	0.00%	0.00%	1.49%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.25%	0.00%	0.00%	0.18%	0.19%
30	1.06%	0.00%	42.86%	3.77%	0.00%	40.00%	0.00%	31.33%	124.49%	0.00%	7.40%	0.00%	0.00%	2.67%	0.41%
31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.88%	0.00%	0.00%	0.22%	0.14%
33	2.87%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.92%	1.00%	0.00%	0.00%	0.11%	0.19%
34	1.59%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.17%	0.00%	0.00%	0.00%	0.30%	0.21%
35	0.84%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.11%	0.17%
36	1.29%	0.00%	3.37%	0.00%	0.00%	1.05%	0.00%	0.00%	0.00%	0.00%	0.48%	0.00%	0.00%	0.16%	57.30%
37	83.59%	0.00%	272.34%	9.28%	0.00%	31.82%	0.00%	29.27%	122.22%	155.56%	30.77%	0.00%	0.00%	1.78%	0.38%
38	2.53%	0.00%	4.17%	1.49%	0.00%	0.00%	0.00%	2.38%	2.70%	24.14%	1.56%	0.00%	0.00%	0.13%	0.18%
39	16.60%	0.00%	16.05%	8.20%	0.00%	26.36%	0.00%	5.56%	69.23%	8.70%	6.91%	0.00%	0.00%	2.57%	0.67%
40	2.13%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.56%	0.00%	0.00%	0.12%	0.14%
41	0.00%	0.00%	8.16%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.77%	0.00%	0.00%	0.02%	0.19%
42	0.00%	0.00%	7.69%	0.00%	0.00%	7.14%	0.00%	8.93%	17.02%	6.98%	1.39%	0.00%	0.00%	0.32%	0.28%
43	79.49%	0.00%	292.41%	30.77%	0.00%	12.05%	0.00%	34.15%	109.09%	323.08%	35.06%	0.00%	0.00%	2.17%	1.31%
44	2.04%	0.00%	2.41%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.70%	0.80%	0.00%	0.00%	0.66%	0.06%
45	0.56%	0.00%	0.00%	1.52%	0.00%	0.00%	0.00%	2.27%	0.00%	0.00%	1.41%	0.00%	0.00%	0.14%	0.16%
47	1.77%	0.00%	8.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.08%	0.00%	0.00%	0.13%	0.22%
48	60.15%	0.00%	7.53%	5.00%	0.00%	21.13%	0.00%	1.69%	20.00%	167.65%	0.50%	0.00%	0.00%	1.24%	0.39%
49	4.78%	0.00%	1.52%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.12%	0.31%
50	2.21%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.48%	0.00%	0.00%	0.04%	0.33%
Arithmetic Mean	11.59%		50.35%	3.07%		6.46%		5.65%	19.85%	36.09%	4.74%			0.50%	1.57%
Median	1.28%		2.82%	0.00%		0.00%		0.00%	0.00%	0.00%	0.92%			0.13%	0.21%

Table 3.43: Percentage errors of individual parameter fits attained from simultaneously fitting the full patient population $\{\widehat{P}_1, ..., \widehat{P}_{50}\}$ generated from the LNP model while only observing A and I data. Here the badly fitted patients $\widehat{P}_{18}$, $\widehat{P}_{21}$, $\widehat{P}_{22}$, $\widehat{P}_{23}$, $\widehat{P}_{32}$, and $\widehat{P}_{46}$ from Table 3.42 are removed from the population prior to fitting. Also shown are the mean and median errors for each parameter. In this case $\mu_{VT} = 15.1$, $\mu_{TI} = 14.19$, $\gamma_T = 0.08$, $\beta_{IB} = 0.0025$, and $s_I = 1000$ are fixed to its exact value for the whole population. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%.

3.6 Reduction of data points

We now discuss the impact of reducing the data resolution on the practical identifiability of model parameters. In our testing of the LNP model thus far, we simulated for each virtual patient a set of 101 data points in the time point vector **t**; the first point being at $t = 0$ days and the remaining one hundred being logarithmically spaced from $t = 0.001$ to $t = 365$ days. This was because our focus has mostly been on confirming theoretical identifiability with a robust data resolution. However, physical patient data used in clinical trails can be quite sparse. To better gauge the effectiveness of the parameter fits in a scenario where real data is used, we reduce the patient datasets from 101 data points to 11 data points. While this is still larger than many clinical trials, it represents a large increase in sparsity. We modify **t** by maintaining the time point at $t = 0$ while reducing the logarithmic set resolution by 90% so that there are only 10 logarithmically spaced data points from $t = 0.001$ to $t = 365$ days. We also maintain the repeated datasets for each patient discussed in Section 2.7 which gave improved parameter fits under SAEM. Hence, we regenerate the patient population $\{\mathcal{P}_1, \ldots, \mathcal{P}_{50}\}$ given the MATLAB generated parameters in Table 2.22 but with 11 data points instead of 101. We write this reduced dataset patient population as $\{\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_{50}\}$ with the same $\mathbf{p}_i$'s as before to refer to the parameter vectors and with $\overline{\mathcal{U}}_i$ referring to the new simulated datasets with 11 data points repeated five times. For consistency, we summarise the generation procedure once again

1. For each $\mathbf{p}_i$ generate the solution matrix U_i' using the `ode45` solver given the model equations
2. The matrices U_i' use the adaptive time stepping scheme and so regenerate them using `spline` given the timespan vector **t** consisting of one time point at $t = 0$ and 10 logarithmically spaced time points from $t = 0.001$ to $t = 365$ which gives the matrices
   $\overline{U}_i = \begin{pmatrix} \mathbf{L} & \mathbf{V} & \mathbf{T} & \mathbf{B} & \mathbf{A} & \mathbf{I} \end{pmatrix}$
3. For each $\overline{U}_i$, define $\overline{\mathcal{U}}_i = \begin{pmatrix} \overline{U}_i & \overline{U}_i & \overline{U}_i & \overline{U}_i & \overline{U}_i \end{pmatrix}^T$
4. Group the parameter vectors and splined solution matrices together as $\overline{\mathcal{P}}_i = \{\mathbf{p}_i, \overline{\mathcal{U}}_i\}$ yielding the virtual patients $\{\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_{50}\}$

Like the previous section, for this regenerated patient population $\{\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_{50}\}$, we set A and I as observed state variables and repeat the parameter fits from Table 3.23 using fixed parameter set $\{\mu_{VT}, \mu_{TI}, \gamma_T, \beta_{IB}, s_I, A_0, I_0\}$ giving the following experimental procedure

1. Load the solution dataset $\overline{U}_i$
2. Set A, and I as observed state variables
3. Load the LNP model equations
4. Fix μ_{VT}, μ_{TI}, γ_T, β_{IB}, s_I, A_0, and I_0 to their exact values from MATLAB with random effects disabled
5. Set all other parameters as variable and to be fitted under the maximum likelihood estimation using SAEM (random effects enabled) with initial values equal to the expected value of the respective uniform distribution
6. Run the auto–init procedure iteratively until the variable initial parameter values have stabilised to a certain neighbourhood in the parameter space
7. Set posterior distributions for data on all compartments to normal distributions (Monolix default)
8. Run the SAEM algorithm which outputs the population parameter fits

The results are shown in Table 3.44. Unsurprisingly, with such a reduced dataset, we observe much higher errors across all fitted parameters. The mean error is 37.04% which is substantially higher than the mean error of 2.47% seen in Table 3.23. Still, μ_{BA} and γ_A have some of the lowest errors with respect to the other parameters which at least confirms that even with a reduced dataset the antibody data can be relatively well fitted if the compartment is observed. We note that the fitted values of μ_{TB} for $\mathcal{P}_{41}$ and $\mathcal{P}_{46}$ were 5.3×10^{-11} and 1.6×10^{-8} respectively. Recalling the percentage error formula from equation 2.8, we see that as the fitted value p_{f} approaches zero, the ratio approaches one. This is exactly the case with these two patients where the fitted values are negligible compared to the exact values and so the errors end up equating to 100%. Given that the highest mean error is for μ_{TB}, we conclude this section by suggesting that, in addition to the implied importance of attaining T–cell and interleukin data from Sections 3.1 and 3.2, data from plasma B–cells becomes important to attain when the number of data points is reduced.

To conclude this chapter, we have demonstrated that if the knowledge of certain population parameters is sufficient enough for them to be fixed to a particular value, then the SAEM algorithm can effectively identify the LNP model parameters even when limited to just antibodies (A) and interleukin (I) as observed state variables. Among these parameters, the identifiability analysis of DAISY in Tables 3.4 and 3.17 implies little to no identifiability for T–cells (T) and limited identifiability for interleukin. This unsurprisingly is reflected by the parameters for those compartments

being mostly fixed in Tables 3.11, 3.24, 3.42, and 3.43. In the final chapter, we will now discuss conclusions for the overall study linking back to the discussions of parameter identifiability given clinical trial data in Chapter 1.

Patient ID \ Parameter	μ_{LV}	μ_{VT}	μ_{TB}	μ_{BA}	μ_{TI}	γ_V	γ_T	γ_B	γ_A	γ_I	β_{BI}	β_{IB}	s_I	A_0	I_0
1	1.44%	0.00%	157.14%	20.37%	0.00%	22.50%	0.00%	57.63%	38.64%	0.00%	58.70%	0.00%	0.00%	0.00%	0.00%
2	11.30%	0.00%	7.69%	0.00%	0.00%	8.62%	0.00%	4.17%	14.29%	91.15%	1.30%	0.00%	0.00%	0.00%	0.00%
3	5.69%	0.00%	120.59%	1.96%	0.00%	25.38%	0.00%	32.08%	28.00%	4.35%	30.87%	0.00%	0.00%	0.00%	0.00%
4	10.76%	0.00%	55.00%	8.33%	0.00%	12.00%	0.00%	27.45%	4.00%	51.22%	32.82%	0.00%	0.00%	0.00%	0.00%
5	1498.44%	0.00%	45.00%	1.85%	0.00%	3.57%	0.00%	4.17%	7.69%	251.35%	1.06%	0.00%	0.00%	0.00%	0.00%
6	0.56%	0.00%	7.69%	5.36%	0.00%	2.67%	0.00%	3.19%	0.00%	3.70%	4.46%	0.00%	0.00%	0.00%	0.00%
7	3.37%	0.00%	2.00%	0.00%	0.00%	1.43%	0.00%	0.00%	2.27%	6.90%	0.00%	0.00%	0.00%	0.00%	0.00%
8	3.33%	0.00%	56.67%	6.76%	0.00%	9.09%	0.00%	24.24%	0.00%	37.04%	23.62%	0.00%	0.00%	0.00%	0.00%
9	1.02%	0.00%	0.00%	0.00%	0.00%	2.30%	0.00%	0.00%	2.50%	0.00%	0.38%	0.00%	0.00%	0.00%	0.00%
10	0.00%	0.00%	56.58%	0.00%	0.00%	0.00%	0.00%	2.04%	2.56%	5.13%	3.33%	0.00%	0.00%	0.00%	0.00%
11	14.81%	0.00%	42.14%	15.00%	0.00%	14.29%	0.00%	32.26%	1.96%	32.00%	34.22%	0.00%	0.00%	0.00%	0.00%
12	4.96%	0.00%	30.77%	2.70%	0.00%	21.25%	0.00%	7.41%	25.53%	51.28%	6.44%	0.00%	0.00%	0.00%	0.00%
13	15.71%	0.00%	200.00%	92.58%	0.00%	11.82%	0.00%	42.00%	26.83%	7.50%	87.16%	0.00%	0.00%	0.00%	0.00%
14	14.55%	0.00%	99.84%	6.67%	0.00%	1.85%	0.00%	118.18%	6.00%	92.31%	120.08%	0.00%	0.00%	0.00%	0.00%
15	51.16%	0.00%	76.00%	8.33%	0.00%	8.33%	0.00%	35.71%	0.00%	37.50%	36.32%	0.00%	0.00%	0.00%	0.00%
16	38.46%	0.00%	91.96%	5.56%	0.00%	0.00%	0.00%	23.08%	7.50%	106.25%	19.75%	0.00%	0.00%	0.00%	0.00%
17	8.47%	0.00%	99.85%	15.38%	0.00%	10.77%	0.00%	54.24%	7.69%	0.00%	45.68%	0.00%	0.00%	0.00%	0.00%
18	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	51.06%	0.31%	0.00%	0.00%	0.00%	0.00%
19	8.62%	0.00%	35.62%	2.08%	0.00%	7.69%	0.00%	11.11%	4.08%	23.08%	9.86%	0.00%	0.00%	0.00%	0.00%
20	4.17%	0.00%	66.67%	4.71%	0.00%	18.75%	0.00%	7.27%	20.41%	10.53%	6.18%	0.00%	0.00%	0.00%	0.00%
21	2.78%	0.00%	9.72%	6.67%	0.00%	10.00%	0.00%	2.11%	4.26%	0.00%	0.25%	0.00%	0.00%	0.00%	0.00%
22	1.66%	0.00%	20.90%	30.77%	0.00%	14.94%	0.00%	29.63%	2.17%	4.35%	31.33%	0.00%	0.00%	0.00%	0.00%
23	0.57%	0.00%	3.09%	0.00%	0.00%	6.98%	0.00%	8.57%	14.89%	5.88%	1.10%	0.00%	0.00%	0.00%	0.00%
24	42.58%	0.00%	16.67%	4.26%	0.00%	7.69%	0.00%	5.45%	6.67%	22.22%	4.64%	0.00%	0.00%	0.00%	0.00%
25	397.24%	0.00%	58.11%	17.39%	0.00%	8.16%	0.00%	32.26%	0.00%	14.71%	49.26%	0.00%	0.00%	0.00%	0.00%
26	76.87%	0.00%	69.00%	0.00%	0.00%	0.00%	0.00%	5.88%	2.70%	60.00%	6.54%	0.00%	0.00%	0.00%	0.00%
27	2.61%	0.00%	169.23%	0.00%	0.00%	25.00%	0.00%	33.33%	28.57%	11.43%	34.67%	0.00%	0.00%	0.00%	0.00%
28	0.31%	0.00%	0.00%	17.39%	0.00%	23.08%	0.00%	23.08%	9.09%	5.00%	25.00%	0.00%	0.00%	0.00%	0.00%
29	8.28%	0.00%	394.25%	1.85%	0.00%	14.55%	0.00%	12.50%	10.53%	18.92%	15.00%	0.00%	0.00%	0.00%	0.00%
30	41.48%	0.00%	42.27%	16.87%	0.00%	36.15%	0.00%	16.13%	33.33%	36.36%	22.60%	0.00%	0.00%	0.00%	0.00%
31	0.66%	0.00%	14.29%	3.09%	0.00%	10.00%	0.00%	6.67%	10.00%	22.58%	0.64%	0.00%	0.00%	0.00%	0.00%
32	0.94%	0.00%	0.00%	0.00%	0.00%	1.02%	0.00%	0.00%	0.00%	13.16%	0.31%	0.00%	0.00%	0.00%	0.00%
33	10.25%	0.00%	72.41%	10.42%	0.00%	29.00%	0.00%	9.38%	30.00%	81.08%	5.26%	0.00%	0.00%	0.00%	0.00%
34	26.13%	0.00%	187.67%	5.56%	0.00%	11.70%	0.00%	9.30%	5.00%	21.88%	8.39%	0.00%	0.00%	0.00%	0.00%
35	6.67%	0.00%	294.37%	3.08%	0.00%	24.17%	0.00%	22.58%	27.45%	116.13%	20.00%	0.00%	0.00%	0.00%	0.00%
36	5.26%	0.00%	9.09%	12.99%	0.00%	9.09%	0.00%	19.15%	5.26%	14.29%	20.43%	0.00%	0.00%	0.00%	0.00%
37	1.21%	0.00%	82.86%	19.23%	0.00%	13.70%	0.00%	3.30%	2.27%	6.06%	2.52%	0.00%	0.00%	0.00%	0.00%
38	13.82%	0.00%	93.59%	21.92%	0.00%	20.31%	0.00%	5.26%	10.87%	2.04%	0.89%	0.00%	0.00%	0.00%	0.00%
39	3.95%	0.00%	38.46%	5.13%	0.00%	8.33%	0.00%	11.43%	12.20%	3.85%	12.79%	0.00%	0.00%	0.00%	0.00%
40	5.53%	0.00%	123.53%	19.23%	0.00%	7.14%	0.00%	41.07%	15.79%	0.00%	45.54%	0.00%	0.00%	0.00%	0.00%
41	125.86%	0.00%	100.00%	982.61%	0.00%	580.56%	0.00%	32.56%	16.67%	471.43%	9.23%	0.00%	0.00%	0.00%	0.00%
42	14.29%	0.00%	161.54%	11.11%	0.00%	5.13%	0.00%	28.07%	10.87%	0.00%	31.17%	0.00%	0.00%	0.00%	0.00%
43	19.07%	0.00%	69.23%	0.00%	0.00%	7.55%	0.00%	17.50%	8.51%	5.71%	18.99%	0.00%	0.00%	0.00%	0.00%
44	99.33%	0.00%	102.02%	35.00%	0.00%	50.00%	0.00%	155.88%	33.33%	4.00%	93.63%	0.00%	0.00%	0.00%	0.00%
45	3.59%	0.00%	181.82%	8.16%	0.00%	28.57%	0.00%	16.67%	22.50%	6.25%	16.31%	0.00%	0.00%	0.00%	0.00%
46	5.45%	0.00%	100.00%	263.54%	0.00%	200.00%	0.00%	28.57%	37.50%	12.00%	4.62%	0.00%	0.00%	0.00%	0.00%
47	2.19%	0.00%	58.33%	14.06%	0.00%	28.33%	0.00%	15.38%	35.90%	30.95%	2.05%	0.00%	0.00%	0.00%	0.00%
48	4.23%	0.00%	99.99%	154.55%	0.00%	89.66%	0.00%	23.08%	7.69%	18.18%	33.33%	0.00%	0.00%	0.00%	0.00%
49	1.98%	0.00%	114.29%	9.09%	0.00%	32.50%	0.00%	26.09%	44.19%	27.59%	17.96%	0.00%	0.00%	0.00%	0.00%
50	9.18%	0.00%	43.64%	0.00%	0.00%	7.79%	0.00%	16.67%	8.33%	63.89%	13.25%	0.00%	0.00%	0.00%	0.00%
Arithmetic Mean	52.77%		79.63%	37.43%		29.87%		22.88%	13.13%	39.25%	21.41%				
Median	6.18%		62.50%	6.67%		10.38%		16.67%	8.42%	14.50%	14.12%				

Table 3.44: Percentage errors from fitting parameters simultaneously using the SAEM algorithm on the virtual patient population $\{\overline{\mathcal{P}}_1, \ldots, \overline{\mathcal{P}}_{50}\}$ generated from the LNP model while only observing A and I data. Also shown are the mean and median errors for each parameter. In this case, μ_{VT}, μ_{TI}, γ_T, β_{IB}, s_I, A_0, and I_0 being fixed to their exact values. The errors between 0% and 100% are shaded on a gradient scale given by 0% 20% 100%.

Chapter 4

In this chapter, we provide conclusions from the identifiability analysis of the lipid nano–particle (LNP) vaccine model in Chapters 2 and 3. We outline the key changes made to the testing of the model that helped to improve the accuracy of the parameter fits. We also discuss the effectiveness of coupling the differential algebra analysis of DAISY with the non–linear mixed effects (NLME) modelling of Monolix in order to better identify parameters with lower identifiability prior to fitting. From this we highlight the compartments that can be considered more of a priority when it comes to physical data collection in clinical vaccine trials (Section 4.1). We then discuss the implications for future work with regards to parameter identifiability of in–host vaccine models and improving the effectiveness of parameter fitting with sparse datasets (Section 4.2).

4.1 Conclusions of the identifiability analysis

In Section 2.4, for our initial testing of the full LNP model of equations 2.1a–2.1h, we recall that the results of the parameter fits in Monolix showed high errors for most parameters as seen in Table 2.11. From this, we tested fitting each parameter individually in Section 2.5 with Table 2.12 showing substantial improvements in the errors. However, all parameters from the cytotoxic T–cell (C) and interferon–γ (F) compartments as well as the LNP (L) decay rate γ_L still showcased high errors. Eliminating the L decay rate and the non–linear terms from the C and F equations gave the results of Table 2.18 which still showed high errors for all remaining C and F parameters. Hence, we concluded that the small sizes of parameters like μ_{VC}, γ_L, γ_C, β_{CF}, and β_{FC}, from the distributions in Table 2.5 derived from Korosec et al. [27], likely contributed the difficulty in correctly identifying their correct values in Monolix. The impact of such small parameter values on the model dynamics can be insignificant enough to the point where even a highly dense dataset $(10,000$ linearly spaced points) cannot capture them.

We used the existence of these small parameters to justify our reduction of the model by eliminating the C and F compartments due to them being de–coupled from the antibody dynamics allowing us to focus more on that part of the immune response. We acknowledge that this issue may be addressed by further increasing the density of the simulated datasets beyond those that were tested in Sections 2.4 and 2.5; allowing for the effects of such small parameter values on the dynamics to be captured. We note however that this will likely require much more powerful hardware as one of the main reasons why the dataset resolution was reduced in Section 2.6 was the extremely long runtimes of the SAEM algorithm in the initial testing. Of course, as outlined in Section 2.3, the small values taken on by some of the parameters may be the result of them being non–identifiable given the data used in Korosec et al. from which the distributions were derived. Further testing of laboratory data may help to better inform such distributions. We will discuss this more in Section 4.2. Subsequently, a major conclusion from Sections 2.4, 2.5, and 2.6 is that even for globally identifiable mathematical models where all compartments are observed, the density of simulated data required to captured the dynamics of all model parameters can be high enough such that it can only be achieved by more powerful machines.

In Section 2.7, we showed that by considering small populations of identical virtual patients instead of a single individual, we were able to take better advantage of the individual parameter sampling in the SAEM algorithm yielding lower errors in the parameter fits compared to the use of the Nelder–Mead simplex algorithm for individual datasets in Section 2.6. As a result, this method of utilising identical populations was maintained for the remaining experiments. We propose that this method

may prove useful for analysing parameter identifiability more generally in within–host models using NLME modelling where researchers wish to consider individual datasets in particular. Using identical datasets allows for individuals to be considered as a population allowing for better use of NMLE methods.

In Section 2.9, we demonstrated the effectiveness of taking advantage of the coupling of compartments by making use analytic solutions of the LNP model of equations 2.20a–2.20f wherever they were available. Unlike Section 2.7, where several parameters were fixed to yield more accurate fits despite the model being globally identifiable with all compartments observed, only β_{IB} was fixed as a result of fitting the LNP model successively by compartment. We speculated that the difficulty in correctly identifying β_{IB} was due to its small size given the distribution derived from Korosec et al. [27]. The accuracy of the errors given only one parameter was fixed suggests that in future studies researchers should look to apply analytic solutions if they exist and if their corresponding state variables are observed. In this case, the same method could not be applied in Chapter 3 where only the downstream compartments of plasma B–cells (B), antibodies (A), and interleukin (I) were observed and for which analytic solutions were not attainable.

In Chapter 3, by eliminating the observations of the LNP (L), vaccinated cells (V), and T–cells (T), the LNP model was no longer globally identifiable, The DAISY results allowed for the classification of structural identifiability of the model parameters given the definitions of Section 1.2. Because of this, unlike Chapter 2, we were able to inform our choices of fixing parameters prior to any extensive testing in Monolix. Because the fixing of β_{IB} maintained from Section 2.9 meant that the identifiability classifications of Table 3.4, with B, A, and I observed, and Table 3.17, with A and I observed, were the same. From this a key conclusion we attained from both Sections 3.1 and 3.2 is that partially identifiable parameters, being coupled with other parameters in the DAISY solutions, are accurately identified in Monolix under the condition that one or more of said coupled parameters are fixed. Furthermore, the overall impact on improving the errors by fixing the locally, partially, and non–identifiable parameters as implied by DAISY was substantial as seen in Tables 3.11 and 3.24. This demonstrates the effectiveness of informing the fixing of parameters through the differential algebra analysis. Moreover, the increased errors when observing only A in Table 3.30, even with extra fixed parameters, indicates that the inclusion of I as an observed state variable contributes immensely to identifiability. This is further supported by the fact that the employed hardware was unable to derive the DAISY differential algebra solutions when only observing A. We suspect that even with more powerful hardware allowing for the derivation of the DAISY solutions when observing only A, the complexity of the solutions will mandate the fixing of several extra parameters.

From the results of Chapter 3, we suggest the following conclusions for data collection. The fixed parameter sets from the bootstrapping sample that we then applied to the entire patient population fixed both μ_{VT}, which was non–identifiable, and γ_T, which was locally identifiable. As such, both parameters from the T-cell (T) compartment were fixed meaning that all of the subsequent fits in Monolix assumed exact knowledge of the T–cell dynamics per individual. This is unsurprising as the T–cells are what prime the non–linear dynamics and by extension the antibody immune response. This suggests that attaining further data on the T–cell dynamics following mRNA LNP vaccine administration can help improve the identifiability of in–host immune response models. Similarly, the interleukin priming rate μ_{TI} was fixed in both sets applied to the entire patient population giving low overall errors suggesting that data on interleukin growth from stimulation by T–cells and the regulatory process associated with plasma B–cells can improve identifiability in a similar way. Finally, we suggest that the use of simulated data paired with software such as DAISY and Monolix serves as an excellent tool for verifying the impacts of observed state variables on structural identifiability and assisting in the choice of fixing parameters that are not globally identifiable in order to reduce the errors of the parameter estimates from fitting data. On a more general level, in the broader field of mathematical epidemiology and immunology, such analyses of parameter identifiability using simulated data can help to inform the types of data that studies and trials should focus more on collecting when fitting to mathematical models. We consider the results of this study to be an appropriate demonstration of the effectiveness of simulated data in informing such decisions.

4.2 Future work

Several aspects of this study leave gaps which we now propose for potential future research. Firstly, and perhaps most importantly, the derivation of the parameter distributions used to generate virtual patients was based on the results of Korosec et al. [27] analysing laboratory data. As mentioned before, we recognise that the difficulties associated with fitting small parameters in Chapters 2 and 3 may be the result of said parameters not being identifiable from the available laboratory data in Korosec et al. and hence taking on small values in those results which defined the distributions. However, we cannot conclusively say this from the results of this study. Therefore, we propose further testing of laboratory data in order to better inform the derivation of parameter distributions used to simulate the immune response. From this, we also refer back to the full LNP model which included the compartments for cytotoxic T–cells (C) and interferon–γ (F) that were decoupled from the antibody dynamics. We suggest separate testing and modelling of the immune response associated with these compartments by instead eliminating the plasma B–cells (B), antibodies (A),

and interleukin (I). Scaling the model to increase the order of the small parameters associated with C and F, should such sizes be supported by further laboratory data testing, may help to address the difficulties with fitting that were encountered in this study. Furthermore, re–fitting the existing laboratory data by fixing parameters on both the individual and population levels given the identifiability conclusions from Chapter 3 would help to confirm the small parameter values originally being a result of practical non–identifiability and potentially provide better distributions for simulating data.

We briefly comment on the increased errors seen in virtual patients lacking clear antibody peaks referring back in the example of patient $\mathcal{P}_{13}$ in Section 2.7. We propose future testing of simulated data that specifically keeps antibody growth low to confirm if, more generally, the lack of peaks hinders identifiability especially if the observed state variables are limited to the downstream compartments as they were in Chapter 3. This may be supplemented by testing the immune responses of real patients with weakened immune systems to confirm if such weaknesses contribute to reduced antibody stimulation.

On the topic of the antibody dynamics, while Korosec et al. [27] did observe data from interleukin in addition to the antibodies, a clearer picture on the structural identifiability of the LNP model solely under antibody observations would still be beneficial to attain in increasing our overall understanding of identifiability. However as noted before, the employed hardware was not powerful enough for DAISY to derive the differential algebra solutions for the LNP model given A as the sole observed state variable. We hence propose for future research the use of improved hardware that may be able to calculate said solutions such that a complete classification of structural identifiability of the LNP model observing only antibodies can be attained.

From the results of Section 3.5, we showed that, even when observing only antibodies and interleukin, fixing parameters on a population level resulted in low errors for most virtual patients. Hence if enough relevant data can be obtained so as to discern what these population parameters are in reality, especially if there is a low variation across individuals, then theoretically almost all individuals from a population can have their data fitted well. However, without knowing the exact parameter values, it is difficult to say conclusively which individuals are fitted well and which individuals are not. Therefore, we suggest pairing the parameter fits with other identifiability experiments that are independent from the given patient data in order to help verify the accuracy of the parameter fits.

Finally, from Section 3.6, we showed that practical identifiability in the LNP model remains an issue when the number of data points was reduced to represent data more realistically attained

from clinical trials. We suggest a more detailed analysis that considers a gradual reduction of data points to gauge the impact of the errors of specific parameters as well as considering the observations of more compartments for which physical data is feasibly available.

References

[1] R. M. Anderson and R. M. May. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, 1991. ISBN 978-019-85-4599-6. doi:10.1093/oso/9780198545996.001.0001.

[2] M. A. Nowak et al. Antigenic Diversity Thresholds and the Development of AIDS. Science, 254(5034):963–969, 1991. doi:10.1126/science.1683006.

[3] A. S. Perelson, D. E. Kirschner, and R. De Boer. Dynamics of HIV infection of CD4+ T cells. Mathematical Biosciences, 114(1):81–125, 1993. doi:10.1016/0025-5564(93)90043-a.

[4] D. D. Ho et al. Rapid turnover of plasma virions and CD4 lymphocytes in HIV–1 infection. Nature, 373(6510):123–126, 1995. doi:10.1038/373123a0.

[5] A. S. Perelson et al. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life–Span, and Viral Generation Time. Science, 271(5255):1582–1586, 1996. doi:10.1126/science.271.5255.1582.

[6] M. A. Nowak and C. R. M. Bangham. Population Dynamics of Immune Responses to Persistent Viruses. Science, 272(5258):74–79, 1996. doi:10.1126/science.272.5258.74.

[7] A. V. Herz et al. Viral dynamics in vivo: Limitations on estimates of intracellular delay and virus decay. Proceedings of the National Academy of Sciences, 93(14):7247–7251, 1996. doi:10.1073/pnas.93.14.7247.

[8] A. U. Neumann et al. Hepatitis C Viral Dynamics in vivo and the Antiviral Efficacy of Interferon–α Therapy. Science, 282(5386):103–107, 1998. doi:10.1126/science.282.5386.103.

[9] S. M. Ciupe et al. Modeling the mechanisms of acute hepatitis B virus infection. Journal of Theoretical Biology, 247(1):23—-35, 2007. doi:10.1016/j.jtbi.2007.02.017.

[10] S. M. Ciupe et al. The role of cells refractory to productive infection in acute hepatitis B viral dynamics. Proceedings of the National Academy of Sciences, 104(12):5050–5055, 2007. doi:10.1073/pnas.0603626104.

[11] J. M. Heffernan and M. J. Keeling. Implications of vaccination and waning immunity. Proceedings of the Royal Society B Biological Sciences, 276(1664):2071–2080, 2009. doi:10.1098/rspb.2009.0057.

[12] J. Guedj et al. Modeling shows that the NS5A inhibitor daclatasvir has two modes of action and yields a shorter estimate of the hepatitis C virus half-life. Proceeedings of the National Academy of Sciences, 110(10):3991–3996, 2013. doi:10.1073/pnas.1203110110.

[13] S. M. Ciupe and J. M. Heffernan. In–host modeling. Infectious Disease Modelling, 2(2): 188–202, 2017. doi:10.1016/j.idm.2017.04.002.

[14] Q. Li et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus—Infected Pneumonia. New England Journal of Medicine, 382(13):1199–1207, 2020. doi:10.1056/NEJMoa2001316.

[15] S. A. Lauer et al. The Incubation Period of Coronavirus Disease 2019 (COVID–19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine, 172(9):1–7, 2020. doi:10.7326/M20-0504.

[16] X. He et al. Temporal dynamics in viral shedding and transmissibility of COVID–19. Nature Medicine, 26:672–675, 2020. doi:10.1038/s41591-020-0869-5.

[17] N. Néant et al. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. Proceedings of the National Academy of Sciences, 118(8):e2017962118, 2021. doi:10.1073/pnas.2017962118.

[18] J. A. Wolff et al. Direct Gene Transfer into Mouse Muscle in Vivo. Science, 247(4949): 1465–1468, 1990. doi:10.1126/science.1690918.

[19] F. Martinon et al. Induction of virus-specific cytotoxic T lymphocytes in vivo by liposome-entrapped mRNA. Mathematical Biosciences, 23(7):1719–1722, 1993. doi:10.1002/eji.1830230749.

[20] N. Pardi et al. Expression kinetics of nucleoside-modified mRNA delivered in lipid nanoparticles to mice by various routes. Journal of Controlled Release, 217:345–351, 2015. doi:10.1016/j.jconrel.2015.08.007.

[21] C. Zhang et al. Advances in mRNA Vaccines for Infectious Diseases. Frontiers in Immunology, 10:594, 2019. doi:10.3389/fimmu.2019.00594.

[22] F. Krammer. SARS–CoV–2 vaccines in development. Nature, 586:516–527, 2020. doi:10.1038/s41586-020-2798-3.

[23] U. Sahin et al. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. Nature, 586:594–599, 2020. doi:10.1038/s41586-020-2814-7.

[24] M. W. Tan et al. Differences in virus and immune dynamics for SARS–CoV–2 Delta and Omicron infections by age and vaccination histories. BMC Infectious Diseases, 24:654, 2024. doi:10.1186/s12879-024-09572-x.

[25] Z. Xu et al. A mathematical model simulating the adaptive immune response in various vaccines and vaccination strategies. Scientific Reports, 14:23995, 2024. doi:10.1038/s41598-024-74221-x.

[26] S. Farhang-Sardroodi et al. Analysis of host immunological response of adenovirus–based COVID–19 vaccines. Vaccines, 9(8):861, 2021. doi:10.3390/vaccines9080861.

[27] C. S. Korosec et al. Long-term durability of immune responses to the BNT162b2 and mRNA–1273 vaccines based on dosage, age and sex. Scientific Reports, 12:21232, 2022. doi:10.1038/s41598-022-25134-0.

[28] R. R. Goel et al. Distinct antibody and memory B cell responses in SARS–CoV–2 naïve and recovered individuals after mRNA vaccination. Science Immunology, 6(58):1–13, 2021. doi:10.1126/sciimmunol.abi6950.

[29] M. V. Stankov et al. Humoral and Cellular Immune Responses Against Severe Acute Respiratory Syndrome Coronavirus 2 Variants and Human Coronaviruses After Single BNT162b2 Vaccination. Clinical Infectious Diseases, 73(11):2000–2008, 2021. doi:10.1093/cid/ciab555.

[30] C. Bergamaschi et al. Systemic IL-15, IFN-γ, and IP-10/CXCL10 signature associated with effective immune response to SARS–CoV–2 in BNT162b2 mRNA vaccine recipients. Cell Reports, 36(6):109504, 2021. doi:10.1016/j.celrep.2021.109504.

[31] D. Lozano-Ojalvo et al. Differential effects of the second SARS–CoV–2 mRNA vaccine dose on T cell immunity in naive and COVID–19 recovered individuals. Cell Reports, 36(8): 109570, 2021. doi:10.1016/j.celrep.2021.109570.

[32] A. T. Widge et al. Durability of responses after SARS–CoV–2 mRNA–1273 Vaccination. The New England Journal of Medicine, 384(1):80–82, 2021. doi:10.1056/NEJMc2032195.

[33] Z. Wang et al. mRNA vaccine–elicited antibodies to SARS–CoV–2 and circulating variants. Nature, 592:616–622, 2021. doi:10.1038/s41586-021-03324-6.

[34] M. S. Suthar et al. Durability of immune responses to the BNT162b2 mRNA vaccine. Med, 3(1):25–27, 2021. doi:10.1016/j.medj.2021.12.005.

[35] I. R. Moyles, C. S. Korosec, and J. M. Heffernan. Determination of significant immunological timescales from mRNA–LNP–based vaccines in humans. Journal of Mathematical Biology, 86:86, 2023. doi:10.1007/s00285-023-01919-3.

[36] J. A. DiMasi, R. W. Hansen, and H. G. Grabowski. The price of innovation: new estimates of drug development costs. Journal of Health Economics, 22(2):151—-185, 2003. doi:10.1016/S0167-6296(02)00126-1.

[37] H. Moses et al. Financial Anatomy of Biomedical Research. JAMA, 294(11):1333—-1342, 2005. doi:10.1001/jama.294.11.1333.

[38] D. Kyte et al. Inconsistencies in Quality of Life Data Collection in Clinical Trials: A Potential Source of Bias? Interviews with Research Nurses and Trialists. PLOS One, 8(10):e76625, 2013. doi:10.1371/journal.pone.0076625.

[39] D. Hind et al. Comparative costs and activity from a sample of UK clinical trials units. Trials, 18:203, 2017. doi:10.1186/s13063-017-1934-3.

[40] M. C. Eisenberg and H. V. Jain. A confidence building exercise in data and identifiability: Modeling cancer chemotherapy as a case study. Journal of Theoretical Biology, 431:63–78, 2017. doi:10.1016/j.jtbi.2017.07.018.

[41] K. Roosa and G. Chowell. Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models. Theoretical Biology and Medical Modelling, 16:1, 2019. doi:10.1186/s12976-018-0097-6.

[42] O-T. Chis, J. R. Banga, and E. Balsa-Canto. Structural Identifiability of Systems Biology Models: A Critical Comparison of Methods. PLoS ONE, 6(11):e27755, 2011. doi:10.1371/journal.pone.0027755.

[43] D. J. Cole. Parameter Redundancy and Identifiability. Chapman and Hall / CRC, New York, 1st edition, 2020. ISBN 978-131-51-2000-3.

[44] N. Cunniffe et al. Identifiability and Observability in Epidemiological Models. Springer Singapore, Singapore, 1st edition, 2024. ISBN 978-981-97-2539-7.

[45] M. P. Saccomani et al. DAISY. University of Padova and University of Cagliari. URL `https://daisy.dei.unipd.it/`.

[46] S. Audoly et al. Global identifiability of nonlinear models of biological systems. IEEE Transactions on Biomedical Engineering, 48(1):55–65, 2001. doi:10.1109/10.900248.

[47] M. P. Saccomani, S. Audoly, and L. D'Angiò. Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica, 39(4):619–632, 2003. doi:10.1016/S0005-1098(02)00302-3.

[48] G. Bellu et al. DAISY: A new software tool to test global identifiability of biological and physiological systems. Computer Methods and Programs in Biomedicine, 88(1):52–61, 2007. doi:10.1016/j.cmpb.2007.07.002.

[49] M. P. Saccomani et al. Examples of testing global identifiability of biological and biomedical models with the DAISY software. Computers in Biology and Medicine, 40(4):402–407, 2010. doi:10.1016/j.compbiomed.2010.02.004.

[50] A. Raue et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, 25(15):1923–1929, 2009. doi:10.1093/bioinformatics/btp358.

[51] F-G. Wieland et al. On structural and practical identifiability. Current Opinion in Systems Biology, 25:60–69, 2021. doi:10.1016/j.coisb.2021.03.005.

[52] M. Heinrich et al. On structural and practical identifiability: Current status and update of results. Current Opinion in Systems Biology, 41:100546, 2025. doi:10.1016/j.coisb.2025.100546.

[53] M. Lavielle. Mixed Effects Models for the Population Approach: Models, Tasks, Methods, and Tools. Chapman and Hall/CRC, New York, 1st edition, 2014. ISBN 978-042-91-5718-9. doi:10.1201/b17203.

[54] Lixoft. Monolix, . URL `https://lixoft.com/products/monolix/`.

[55] Lixoft. Monolix Documentation. Lixoft (A Simulations Plus Company), . URL `https://monolixsuite.slp-software.com/monolix/2024R1/`.

[56] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. Annals of Statistics, 27(1):94–128, 1999. doi:10.1214/aos/1018031103.

[57] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis, 49(4):1020–1038, 2005. doi:10.1016/j.csda.2004.07.002.

[58] Lixoft. Monolix Documentation: Loading a Data Set. Lixoft (A Simulations Plus Company), . URL `https://monolixsuite.slp-software.com/monolix/2024R1/loading-a-data-set`.

[59] Lixoft. Monolix Documentation: Setting the Structural Model. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/setting-the-structural-model.

[60] Lixoft. Monolix Documentation: Initial Estimates. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/initial-estimates.

[61] Lixoft. Monolix Documentation: SAEM. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/saem.

[62] Lixoft. Monolix Documentation: Bayesian Estimation. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/bayesian-estimation-or-fixed-population-parameters.

[63] J. C. Lagarias et al. Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions. SIAM Journal on Optimization, 9(1):112–147, 1998. doi:10.1137/S1052623496303470.

[64] Lixoft. Monolix Documentation: Check initial estimates and auto–init. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/check-initial-estimates-and-auto-init.

[65] Lixoft. Monolix Documentation: Observation Model. Lixoft (A Simulations Plus Company), . URL https://monolixsuite.slp-software.com/monolix/2024R1/observation-error-model.

[66] MathWorks. MATLAB Documentation: fminsearch. MathWorks, . URL https://www.mathworks.com/help/matlab/ref/fminsearch.html.

[67] P. Traynard et al. Efficient Pharmacokinetic Modeling Workflow With the MonolixSuite: A Case Study of Remifentanil. CPT: Pharmacometrics & Systems Pharmacology, 9(4):198–210, 2020. doi:10.1002/psp4.12500.

[68] N. Heitzman-Breen et al. The effect of model structure and data availability on Usutu virus dynamics at three biological scales. Royal Society Open Science, 11:231146, 2024. doi:10.1098/rsos.231146.

[69] MathWorks. MATLAB Documentation: rand. MathWorks, . URL https://www.mathworks.com/help/matlab/ref/double.rand.html.

[70] MathWorks. MATLAB Documentation: ode45. MathWorks, . URL https://www.mathworks.com/help/matlab/ref/ode45.html.

[71] J. R. Dormand and P. J. Prince. A family of embedded Runge–Kutta formulae. Journal of Computational and Applied Mathematics, 6(1):19–26, 1980. doi:10.1016/0771-050X(80)90013-3.

[72] L. F. Shampine and M. W. Reichelt. The MATLAB ODE suite. SIAM Journal on Scientific Computing, 18(1):1–22, 1997. doi:10.1137/S1064827594276424.

[73] MathWorks. MATLAB Documentation: spline. MathWorks, . URL https://www.mathworks.com/help/matlab/ref/spline.html.

[74] K. Eriksson, C. Johnson, and A. Logg. Explicit Time–Stepping for Stiff ODEs. SIAM Journal on Scientific Computing, 25(4):1142—-1157, 2004. doi:10.1137/S1064827502409626.

[75] P. Bocher et al. Explicit Runge—Kutta Methods for Stiff Problems with a Gap in Their Eigenvalue Spectrum. Journal of Scientific Computing, 77:1055–1083, 2018. doi:https://doi.org/10.1007/s10915-018-0737-3.

[76] MathWorks. MATLAB Documentation: ode23t. MathWorks, . URL https://www.mathworks.com/help/matlab/ref/ode23t.html.

[77] L. F. Shampine, M. W. Reichelt, and J. A. Kierzenka. Solving Index–1 DAEs in MATLAB and Simulink. SIAM Review, 41(3):538–552, 1999. doi:10.1137/S003614459933425X.

[78] M. E. Hosea and L. F. Shampine. Analysis and implementation of TR–BDF2. Applied Numerical Mathematics, 22(1–2):21–37, 1996. doi:10.1016/0168-9274(95)00115-8.

Appendix A

Derivations of analytic solutions and algebraic manipulations

2.1

L, V, T, A solutions from the full LNP model in equations 2.1a – 2.1h. Starting with L, we have

$$\frac{dL}{dt} = -\mu_{LV}L - \gamma_L L = -(\mu_{LV} + \gamma_L)L$$

$$\int \frac{1}{L}\, dL = \int -(\mu_{LV} + \gamma_L)\, dt$$

$$\ln L = -(\mu_{LV} + \gamma_L)t + \mathcal{C}$$

$$L = e^{-(\mu_{LV}+\gamma_L)t+\mathcal{C}}$$

$$L(0) = e^{\mathcal{C}} = 1$$

$$\mathcal{C} = 0$$

$$L(t) = e^{-(\mu_{LV}+\gamma_L)t}$$

This now allows us to re–write and solve for V

$$\frac{dV}{dt} = \mu_{LV}L - \gamma_V V = \mu_{LV}e^{-(\mu_{LV}+\gamma_L)t} - \gamma_V V$$

$$\frac{dV}{dt} + \gamma_V V = \mu_{LV}e^{-(\mu_{LV}+\gamma_L)t}$$

$$\exp\left[\int \gamma_V\, dt\right] = e^{\gamma_V t} \quad \text{(integrating factor)}$$

$$e^{\gamma_V t}\frac{dV}{dt} + \gamma_V e^{\gamma_V t} V = \frac{d}{dt}\left[e^{\gamma_V t} V\right] = \mu_{LV} e^{(\gamma_V - \mu_{LV} - \gamma_L)t}$$

$$\int \frac{d}{dt}\left[e^{\gamma_V t} V\right]\, dt = \int \mu_{LV} e^{(\gamma_V - \mu_{LV} - \gamma_L)t}\, dt$$

$$e^{\gamma_V t} V = \frac{\mu_{LV} e^{(\gamma_V - \mu_{LV} - \gamma_L)t}}{\gamma_V - \mu_{LV} - \gamma_L} + \mathcal{C}$$

$$V = \frac{\mu_{LV} e^{-(\mu_{LV} + \gamma_L)t}}{\gamma_V - \mu_{LV} - \gamma_L} + \mathcal{C} e^{-\gamma_V t}$$

$$V(0) = \frac{\mu_{LV}}{\gamma_V - \mu_{LV} - \gamma_L} + \mathcal{C} = 0$$

$$\mathcal{C} = -\frac{\mu_{LV}}{\gamma_V - \mu_{LV} - \gamma_L}$$

$$V(t) = \frac{\mu_{LV} e^{-(\mu_{LV} + \gamma_L)t}}{\gamma_V - \mu_{LV} - \gamma_L} - \frac{\mu_{LV} e^{-\gamma_V t}}{\gamma_V - \mu_{LV} - \gamma_L} = \frac{\mu_{LV}(e^{-(\mu_{LV} + \gamma_L)t} - e^{-\gamma_V t})}{\gamma_V - \mu_{LV} - \gamma_L}$$

We solve for T in the same way

$$\frac{dT}{dt} = \mu_{VT} V - \gamma_T T = \frac{\mu_{VT}\mu_{LV}(e^{-(\mu_{LV} + \gamma_L)t} - e^{-\gamma_V t})}{\gamma_V - \mu_{LV} - \gamma_L} - \gamma_T T$$

$$\frac{dT}{dt} + \gamma_T T = \frac{\mu_{VT}\mu_{LV}(e^{-(\mu_{LV} + \gamma_L)t} - e^{-\gamma_V t})}{\gamma_V - \mu_{LV} - \gamma_L}$$

$$\exp\left[\int \gamma_T\, dt\right] = e^{\gamma_T t} \quad \text{(integrating factor)}$$

$$e^{\gamma_T t}\frac{dT}{dt} + \gamma_T e^{\gamma_T t} T = \frac{d}{dt}\left[e^{\gamma_T t} T\right] = \frac{\mu_{VT}\mu_{LV}(e^{(\gamma_T - \mu_{LV} - \gamma_L)t} - e^{(\gamma_T - \gamma_V)t})}{\gamma_V - \mu_{LV} - \gamma_L}$$

$$\int \frac{d}{dt}\left[e^{\gamma_T t}T\right]\,dt = \int \frac{\mu_{VT}\mu_{LV}(e^{(\gamma_T-\mu_{LV}-\gamma_L)t} - e^{(\gamma_T-\gamma_V)t})}{\gamma_V-\mu_{LV}-\gamma_L}\,dt$$

$$e^{\gamma_T t}T = \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{e^{(\gamma_T-\mu_{LV}-\gamma_L)t}}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{e^{(\gamma_T-\gamma_V)t}}{\gamma_T-\gamma_V}\right) + \mathcal{C}$$

$$T = \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{e^{-(\mu_{LV}+\gamma_L)t}}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{e^{-\gamma_V t}}{\gamma_T-\gamma_V}\right) + \mathcal{C}e^{-\gamma_T t}$$

$$T(0) = \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{1}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{1}{\gamma_T-\gamma_V}\right) + \mathcal{C} = 0$$

$$\mathcal{C} = -\frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{1}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{1}{\gamma_T-\gamma_V}\right)$$

$$T(t) = \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{e^{-(\mu_{LV}+\gamma_L)t}}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{e^{-\gamma_V t}}{\gamma_T-\gamma_V}\right) - \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{e^{-\gamma_T t}}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{e^{-\gamma_T t}}{\gamma_T-\gamma_V}\right)$$

$$= \frac{\mu_{VT}\mu_{LV}}{\gamma_V-\mu_{LV}-\gamma_L}\left(\frac{e^{-(\mu_{LV}+\gamma_L)t} - e^{-\gamma_T t}}{\gamma_T-\mu_{LV}-\gamma_L} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T-\gamma_V}\right)$$

While dB/dt is non–linear and cannot be solved analytically, dA/dt solely depends on B and is of the same linear form as those solved thus far. Therefore, the solution for A can be written in a compact integral form given in terms of the unknown solution $B(t)$

$$\frac{dA}{dt} = \mu_{BA}B(t) - \gamma_A A$$

$$\frac{dA}{dt} + \gamma_A A = \mu_{BA}B(t)$$

$$\exp\left[\int \gamma_A\,dt\right] = e^{\gamma_A t} \quad \text{(integrating factor)}$$

$$e^{\gamma_A t}\frac{dA}{dt} + \gamma_A e^{\gamma_A t}A = \frac{d}{dt}\left[e^{\gamma_A t}A\right] = \mu_{BA}e^{\gamma_A t}B(t)$$

$$\int \frac{d}{dt}\left[e^{\gamma_A t}A\right]\,dt = \int \mu_{BA}e^{\gamma_A t}B(t)\,dt$$

$$e^{\gamma_A t}A = \mu_{BA}\int e^{\gamma_A t}B(t)\,dt + \mathcal{C}$$

$$\text{Let} \quad H(t) = \int e^{\gamma_A t}B(t)\,dt$$

$$A = \mu_{BA}e^{-\gamma_A t}\int e^{\gamma_A t}B(t)\,dt + \mathcal{C}e^{-\gamma_A t} = \mu_{BA}e^{-\gamma_A t}H(t) + \mathcal{C}e^{-\gamma_A t}$$

$$A(0) = \mu_{BA}H(0) + \mathcal{C} = A_0$$

$$\mathcal{C} = A_0 - \mu_{BA}H(0)$$

$$A(t) = \mu_{BA}e^{-\gamma_A t}H(t) + (A_0 - \mu_{BA}H(0))e^{-\gamma_A t}$$

$$= e^{-\gamma_A t}\left[A_0 + \mu_{BA}(H(t) - H(0))\right]$$

$$= e^{-\gamma_A t}\left[A_0 + \mu_{BA}\int_0^t e^{\gamma_A s}B(s)\,ds\right]$$

2.2

Algebraic manipulations to show structural identifiability for the multiple parameter coefficients

$$\mu_{LV} + \gamma_L = \widehat{\mu}_{LV} + \widehat{\gamma}_L$$
$$\mu_{LV} + \gamma_L = \mu_{LV} + \widehat{\gamma}_L$$
$$\gamma_L = \widehat{\gamma}_L$$

$$s_F \gamma_F = \widehat{s}_F \widehat{\gamma}_F$$
$$s_F \gamma_F = s_F \widehat{\gamma}_F$$
$$\gamma_F = \widehat{\gamma}_F$$

$$s_I \gamma_I = \widehat{s}_I \widehat{\gamma}_I$$
$$s_I \gamma_I = s_I \widehat{\gamma}_I$$
$$\gamma_I = \widehat{\gamma}_I$$

$$s_F \beta_{FC} = \widehat{s}_F \widehat{\beta}_{FC}$$
$$s_F \beta_{FC} = s_F \widehat{\beta}_{FC}$$
$$\beta_{FC} = \widehat{\beta}_{FC}$$

$$s_I \beta_{IB} = \widehat{s}_I \widehat{\beta}_{IB}$$
$$s_I \beta_{IB} = s_I \widehat{\beta}_{IB}$$
$$\beta_{IB} = \widehat{\beta}_{IB}$$

2.5

Solution of $\widetilde{F}$ attained by taking equation 2.1g from the full LNP model and setting $\beta_{FC} = 0$ which gives equation 2.17g

$$\frac{d\widetilde{F}}{dt} = \mu_{TF} T - \gamma_F \widetilde{F}$$

We start by substituting the solution for T from above but with $\gamma_L = 0$ which gives

$$\frac{d\widetilde{F}}{dt} = \frac{\mu_{TF}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left(\frac{e^{-\mu_{LV}t} - e^{-\gamma_T t}}{\gamma_T - \mu_{LV}} - \frac{e^{-\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V} \right) - \gamma_F \widetilde{F}$$

For simplicity, let $k = \dfrac{\mu_{TF}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}}$. Then we have

$$\frac{d\widetilde{F}}{dt} + \gamma_F \widetilde{F} = k \left(\frac{e^{\mu_{LV}t} - e^{-\gamma_V t}}{\gamma_T - \mu_{LV}} - \frac{e^{\gamma_V t} - e^{-\gamma_T t}}{\gamma_T - \gamma_V} \right)$$

$$\exp\left[\int \gamma_F \, dt\right] = e^{\gamma_F t} \quad \text{(integrating factor)}$$

$$e^{\gamma_F t}\frac{d\widetilde{F}}{dt} + \gamma_F e^{\gamma_F t}\widetilde{F} = \frac{d}{dt}\left[e^{\gamma_F t}\widetilde{F}\right] = k\left(\frac{e^{(\gamma_F-\mu_{LV})t}}{\gamma_T-\mu_{LV}} - \frac{e^{(\gamma_F-\gamma_V)t}}{\gamma_T-\mu_{LV}} - \frac{e^{(\gamma_F-\gamma_V)t}}{\gamma_T-\gamma_V} + \frac{e^{(\gamma_F-\gamma_T)t}}{\gamma_T-\gamma_V}\right)$$

$$\int \frac{d}{dt}\left[e^{\gamma_F t}\widetilde{F}\right] dt = \int k\left(\frac{e^{(\gamma_F-\mu_{LV})t}}{\gamma_T-\mu_{LV}} - \frac{e^{(\gamma_F-\gamma_V)t}}{\gamma_T-\mu_{LV}} - \frac{e^{(\gamma_F-\gamma_V)t}}{\gamma_T-\gamma_V} + \frac{e^{(\gamma_F-\gamma_T)t}}{\gamma_T-\gamma_V}\right)$$

$$e^{\gamma_F t}\widetilde{F} = k\left(\frac{e^{(\gamma_F-\mu_{LV})t}}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{e^{(\gamma_F-\gamma_V)t}}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{e^{(\gamma_F-\gamma_V)t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{e^{(\gamma_F-\gamma_T)t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right) + \mathcal{C}$$

$$\widetilde{F} = k\left(\frac{e^{-\mu_{LV}t}}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{e^{-\gamma_V t}}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{e^{-\gamma_V t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{e^{-\gamma_T t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right)$$

$$+ \mathcal{C}e^{-\gamma_F t}$$

$$\widetilde{F}(0) = k\left(\frac{1}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{1}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{1}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{1}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right) + \mathcal{C}$$

$$= 0$$

$$\mathcal{C} = -k\left(\frac{1}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{1}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{1}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{1}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right)$$

$$\widetilde{F}(t) = k\left(\frac{e^{-\mu_{LV}t}}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{e^{-\gamma_V t}}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{e^{-\gamma_V t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{e^{-\gamma_T t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right)$$

$$- k\left(\frac{e^{-\gamma_F t}}{(\gamma_T-\mu_{LV})(\gamma_F-\mu_{LV})} - \frac{e^{-\gamma_F t}}{(\gamma_T-\mu_{LV})(\gamma_F-\gamma_V)} - \frac{e^{-\gamma_F t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_V)} + \frac{e^{-\gamma_F t}}{(\gamma_T-\gamma_V)(\gamma_F-\gamma_T)}\right)$$

$$= \frac{\mu_{TF}\mu_{VT}\mu_{LV}}{\gamma_V - \mu_{LV}} \left(\frac{e^{\mu_{LV}t} - e^{-\gamma_F t}}{(\gamma_T - \mu_{LV})(\gamma_F - \mu_{LV})} - \frac{e^{-\gamma_V t - \gamma_F t}}{(\gamma_T - \mu_{LV})(\gamma_F - \gamma_V)} - \frac{e^{-\gamma_V t - \gamma_F t}}{(\gamma_T - \gamma_V)(\gamma_F - \gamma_V)} + \frac{e^{-\gamma_T t - \gamma_F t}}{(\gamma_T - \gamma_V)(\gamma_F - \gamma_T)} \right)$$

2.8

The solutions of L, V, T for the LNP model in equations 2.20a – 2.20c are the same as those derived above for Section 2.1 but with $\gamma_L = 0$.

Considering the solutions of V and T when $\gamma_V = \gamma_T = \gamma$. For V we only need to change the decay parameter to γ

$$V(t) = \frac{\mu_{LV}}{\gamma - \mu_{LV}}(e^{-\mu_{LV}t} - e^{-\gamma t})$$

We can now solve for T

$$\frac{dT}{dt} + \gamma T = \frac{\mu_{VT}\mu_{LV}}{\gamma - \mu_{LV}}(e^{-\mu_{LV}t} - e^{-\gamma t})$$

$$\exp\left[\int \gamma \, dt\right] = e^{\gamma t} \quad \text{(integrating factor)}$$

$$e^{\gamma t}\frac{dT}{dt} + \gamma e^{\gamma t} T = \frac{d}{dt}\left[e^{\gamma t} T\right] = \frac{\mu_{VT}\mu_{LV}}{\gamma - \mu_{LV}}(e^{(\gamma - \mu_{LV})t} - 1)$$

$$\int \frac{d}{dt}\left[e^{\gamma t} T\right] \, dt = \int \frac{\mu_{VT}\mu_{LV}}{\gamma - \mu_{LV}}(e^{(\gamma - \mu_{LV})t} - 1) \, dt$$

$$e^{\gamma t} T = \frac{\mu_{VT}\mu_{LV}}{\gamma - \mu_{LV}} \left(\frac{e^{(\gamma - \mu_{LV})t}}{\gamma - \mu_{LV}} - t \right) + \mathcal{C}$$

$$T = \frac{\mu_{VT}\mu_{LV} e^{-\gamma t}}{\gamma - \mu_{LV}} \left(\frac{e^{(\gamma - \mu_{LV})t}}{\gamma - \mu_{LV}} - t \right) + \mathcal{C} e^{-\gamma t}$$

$$T(0) = \frac{\mu_{VT}\mu_{LV}}{\gamma - \mu_{LV}} \left(\frac{1}{\gamma - \mu_{LV}} \right) + \mathcal{C} = 0$$

$$\mathcal{C} = \frac{\mu_{VT}\mu_{LV}}{(\gamma - \mu_{LV})^2}$$

$$T(t) = \frac{\mu_{VT}\mu_{LV}e^{-\gamma t}}{\gamma - \mu_{LV}} \left(\frac{e^{(\gamma-\mu_{LV})t}}{\gamma - \mu_{LV}} - t \right) + \frac{\mu_{VT}\mu_{LV}e^{-\gamma t}}{(\gamma - \mu_{LV})^2}$$

$$= \frac{\mu_{VT}\mu_{LV}e^{-\gamma t}}{(\gamma - \mu_{LV})^2}(e^{(\gamma-\mu_{LV})t} - (\gamma - \mu_{LV})t) + \frac{\mu_{VT}\mu_{LV}e^{-\gamma t}}{(\gamma - \mu_{LV})^2}$$

$$= \frac{\mu_{VT}\mu_{LV}e^{-\gamma t}}{(\gamma - \mu_{LV})^2}(e^{(\gamma-\mu_{LV})t} - (\gamma - \mu_{LV})t + 1)$$

$$= \frac{\mu_{VT}\mu_{LV}}{(\gamma - \mu_{LV})^2}(e^{-\mu_{LV}t} + e^{-\gamma t} - (\gamma - \mu_{LV})te^{-\gamma t})$$

Appendix B

DAISY differential algebra system output logs

2.2

Given the full LNP model from equations 2.1a – 2.1h while observing all compartments as state variables

```
MODEL EQUATION(S)$

c_ := {df(lx,t)= - lx*mulv,
df(vx,t)= - gammav*vx + lx*mulv,
df(tx,t)= - gammat*tx + muvt*vx,
df(bx,t)=( - (bx*gammab - mutb*tx)*(ix + si) + betabi*bx*ix)/(ix + si),
df(ax,t)= - ax*gammaa + bx*muba,
df(ix,t)= - (gammai*ix - muti*tx) - betaib*bx*ix,
ly=lx,
vy=vx,
ty=tx,
by=bx,
ay=ax,
iy=ix}$

UNKNOWN PARAMETER VECTOR$

b1_ := {mulv,
muvt,
mutb,
muba,
muti,
gammav,
gammat,
gammab,
gammaa,
gammai,
betabi,
betaib,
si}$
```

```
CHARACTERISTIC SET$

aa_(1) := df(ly,t) + ly*mulv$

aa_(2) := df(vy,t) - ly*mulv + vy*gammav$

aa_(3) := df(ty,t) + ty*gammat - vy*muvt$

aa_(4) := df(by,t)*iy + df(by,t)*si + by*iy*( - betabi + gammab) + by*
   gammab*si - iy*ty*mutb - ty*mutb*si$

aa_(5) := df(ay,t) + ay*gammaa - by*muba$

aa_(6) := df(iy,t) + by*iy*betaib + iy*gammai - ty*muti$

aa_(7) :=  - lx + ly$

aa_(8) :=  - vx + vy$

aa_(9) :=  - tx + ty$

aa_(10) :=  - bx + by$

aa_(11) :=  - ax + ay$

aa_(12) :=  - ix + iy$

MODEL ALGEBRAICALLY OBSERVABLE$

PARAMETER VALUES TO SWITCH FROM SYMBOLIC TO NUMERICAL CALCULATIONS$

b2_ := {mulv=2,
muvt=3,
mutb=5,
muba=7,
muti=11,
gammav=13,
```

```
gammat=17,
gammab=19,
gammaa=23,
gammai=29,
betabi=31,
betaib=37,
si=41}$

EXHAUSTIVE SUMMARY (flist)$

flist_ := {mulv - 2,
muvt - 3,
mutb - 5,
muba - 7,
muti - 11,
gammav - 13,
gammat - 17,
gammab - 19,
gammaa - 23,
gammai - 29,
betabi - 31,
betaib - 37,
si - 41}$

MODEL PARAMETER SOLUTION(S)$

g_ := {{betabi=31,
betaib=37,
gammaa=23,
gammab=19,
gammai=29,
gammat=17,
gammav=13,
muba=7,
mulv=2,
mutb=5,
muti=11,
muvt=3,
```

```
si=41}}$

MODEL GLOBALLY IDENTIFIABLE$

INITIAL CONDITION(S) NOT NECESSARY$
```

3.1

Given the LNP model from equations 2.20a – 2.20f while observing *B*, *A*, and *I* as state variables

```
MODEL EQUATION(S)$

c_ := {df(lx,t)= - lx*mulv,
df(vx,t)= - gammav*vx + lx*mulv,
df(tx,t)= - gammat*tx + muvt*vx,
df(bx,t)=( - (bx*gammab - mutb*tx)*(ix + si) + betabi*bx*ix)/(ix + si),
df(ax,t)= - ax*gammaa + bx*muba,
df(ix,t)= - (gammai*ix - muti*tx) - betaib*bx*ix,
by=bx,
ay=ax,
iy=ix}$

UNKNOWN PARAMETER VECTOR$

b1_ := {mulv,
muvt,
mutb,
muba,
muti,
gammav,
gammat,
gammab,
gammaa,
gammai,
betabi,
betaib,
si}$

CHARACTERISTIC SET$
```

```
aa_(1) := df(ay,t) + ay*gammaa - by*muba$

aa_(2) := df(by,t)*iy*muti + df(by,t)*muti*si - df(iy,t)*iy*mutb - df(
   iy,t)*mutb*si - by*iy**2*betaib*mutb +
by*iy*( - betabi*muti - betaib*mutb*si + gammab*muti) + by*gammab*muti*
   si - iy**2*gammai*mutb - iy*gammai*mutb
*si$

aa_(3) := df(by,t,3)*iy**4*betaib*mutb**2 + 3*df(by,t,3)*iy**3*betaib*
   mutb**2*si + 3*df(by,t,3)*iy**2*betaib*
mutb**2*si**2 + df(by,t,3)*iy*betaib*mutb**2*si**3 + 5*df(by,t,2)*df(by
   ,t)*iy**3*betaib*mutb*muti + 15*df(by,t
,2)*df(by,t)*iy**2*betaib*mutb*muti*si + 15*df(by,t,2)*df(by,t)*iy*
   betaib*mutb*muti*si**2 + 5*df(by,t,2)*df(by
,t)*betaib*mutb*muti*si**3 - 3*df(by,t,2)*by*iy**4*betaib**2*mutb**2 +
   df(by,t,2)*by*iy**3*betaib*mutb*( - 3*
betabi*muti - 9*betaib*mutb*si + 3*gammab*muti + gammat*muti + gammav*
   muti) + 3*df(by,t,2)*by*iy**2*betaib*
mutb*si*( - 2*betabi*muti - 3*betaib*mutb*si + 3*gammab*muti + gammat*
   muti + gammav*muti) + 3*df(by,t,2)*by*iy
*betaib*mutb*si**2*( - betabi*muti - betaib*mutb*si + 3*gammab*muti +
   gammat*muti + gammav*muti) + df(by,t,2)*
by*betaib*mutb*muti*si**3*(3*gammab + gammat + gammav) + df(by,t,2)*iy
   **4*betaib*mutb**2*( - 3*gammai + gammat
 + gammav + mulv) + df(by,t,2)*iy**3*mutb*( - 9*betaib*gammai*mutb*si +
     3*betaib*gammat*mutb*si + 3*betaib*
gammav*mutb*si + 3*betaib*mulv*mutb*si + gammai*gammat*muti + gammai*
   gammav*muti + gammat*gammav*muti) + 3*df(
by,t,2)*iy**2*mutb*si*( - 3*betaib*gammai*mutb*si + betaib*gammat*mutb*
   si + betaib*gammav*mutb*si + betaib*
mulv*mutb*si + gammai*gammat*muti + gammai*gammav*muti + gammat*gammav*
   muti) + df(by,t,2)*iy*mutb*si**2*( - 3*
betaib*gammai*mutb*si + betaib*gammat*mutb*si + betaib*gammav*mutb*si +
    betaib*mulv*mutb*si + 3*gammai*gammat*
muti + 3*gammai*gammav*muti + 3*gammat*gammav*muti) + df(by,t,2)*mutb*
   muti*si**3*(gammai*gammat + gammai*
gammav + gammat*gammav) - 2*df(by,t)**2*by*iy**3*betaib**2*mutb*muti -
```

```
   6*df(by,t)**2*by*iy**2*betaib**2*mutb*
muti*si + 2*df(by,t)**2*by*iy*betaib*muti*si*( - betabi*muti - 3*betaib
   *mutb*si) - 2*df(by,t)**2*by*betaib*
muti*si**2*(betabi*muti + betaib*mutb*si) - 2*df(by,t)**2*iy**4*betaib
   **2*mutb**2 + 2*df(by,t)**2*iy**3*betaib
*mutb*( - betabi*muti - 3*betaib*mutb*si + gammab*muti - gammai*muti +
   gammat*muti + gammav*muti + mulv*muti)
+ 2*df(by,t)**2*iy**2*betaib*mutb*si*( - 2*betabi*muti - 3*betaib*mutb*
   si + 3*gammab*muti - 3*gammai*muti + 3*
gammat*muti + 3*gammav*muti + 3*mulv*muti) + 2*df(by,t)**2*iy*betaib*
   mutb*si**2*( - betabi*muti - betaib*mutb*
si + 3*gammab*muti - 3*gammai*muti + 3*gammat*muti + 3*gammav*muti + 3*
   mulv*muti) + 2*df(by,t)**2*betaib*mutb*
muti*si**3*(gammab - gammai + gammat + gammav + mulv) + df(by,t)*df(iy,
   t,2)*iy**3*betaib*mutb**2 + 3*df(by,t)*
df(iy,t,2)*iy**2*betaib*mutb**2*si + 3*df(by,t)*df(iy,t,2)*iy*betaib*
   mutb**2*si**2 + df(by,t)*df(iy,t,2)*
betaib*mutb**2*si**3 + 2*df(by,t)*by**2*iy**4*betaib**3*mutb**2 + df(by
   ,t)*by**2*iy**3*betaib**2*mutb*(2*
betabi*muti + 6*betaib*mutb*si - 2*gammab*muti - gammat*muti - gammav*
   muti) + 3*df(by,t)*by**2*iy**2*betaib**2
*mutb*si*(2*betabi*muti + 2*betaib*mutb*si - 2*gammab*muti - gammat*
   muti - gammav*muti) + df(by,t)*by**2*iy*
betaib*si*(2*betabi**2*muti**2 + 4*betabi*betaib*mutb*muti*si - 2*
   betabi*gammab*muti**2 - betabi*gammat*muti**
2 - betabi*gammav*muti**2 + 2*betaib**2*mutb**2*si**2 - 6*betaib*gammab
   *mutb*muti*si - 3*betaib*gammat*mutb*
muti*si - 3*betaib*gammav*mutb*muti*si) + df(by,t)*by**2*betaib*muti*si
   **2*( - 2*betabi*gammab*muti - betabi*
gammat*muti - betabi*gammav*muti - 2*betaib*gammab*mutb*si - betaib*
   gammat*mutb*si - betaib*gammav*mutb*si) +
df(by,t)*by*iy**4*betaib**2*mutb**2*(4*gammai - 3*gammat - 3*gammav -
   2*mulv) + df(by,t)*by*iy**3*betaib*mutb*
(2*betabi*gammai*muti - 3*betabi*gammat*muti - 3*betabi*gammav*muti -
   2*betabi*mulv*muti + 12*betaib*gammai*
mutb*si - 9*betaib*gammat*mutb*si - 9*betaib*gammav*mutb*si - 6*betaib*
   mulv*mutb*si - 2*gammab*gammai*muti + 3
*gammab*gammat*muti + 3*gammab*gammav*muti + 2*gammab*mulv*muti - 2*
```

```
   gammai*gammat*muti - 2*gammai*gammav*muti
+ gammat*mulv*muti + gammav*mulv*muti) + df(by,t)*by*iy**2*betaib*mutb*
   si*(6*betabi*gammai*muti - 6*betabi*
gammat*muti - 6*betabi*gammav*muti - 4*betabi*mulv*muti + 12*betaib*
   gammai*mutb*si - 9*betaib*gammat*mutb*si -
 9*betaib*gammav*mutb*si - 6*betaib*mulv*mutb*si - 6*gammab*gammai*muti
     + 9*gammab*gammat*muti + 9*gammab*
gammav*muti + 6*gammab*mulv*muti - 6*gammai*gammat*muti - 6*gammai*
   gammav*muti + 3*gammat*mulv*muti + 3*gammav
*mulv*muti) + df(by,t)*by*iy*si*(4*betabi*betaib*gammai*mutb*muti*si -
   3*betabi*betaib*gammat*mutb*muti*si - 3
*betabi*betaib*gammav*mutb*muti*si - 2*betabi*betaib*mulv*mutb*muti*si
   - betabi*gammai*gammat*muti**2 - betabi
*gammai*gammav*muti**2 - betabi*gammat*gammav*muti**2 + 4*betaib**2*
   gammai*mutb**2*si**2 - 3*betaib**2*gammat*
mutb**2*si**2 - 3*betaib**2*gammav*mutb**2*si**2 - 2*betaib**2*mulv*
   mutb**2*si**2 - 6*betaib*gammab*gammai*
mutb*muti*si + 9*betaib*gammab*gammat*mutb*muti*si + 9*betaib*gammab*
   gammav*mutb*muti*si + 6*betaib*gammab*
mulv*mutb*muti*si - 6*betaib*gammai*gammat*mutb*muti*si - 6*betaib*
   gammai*gammav*mutb*muti*si + 3*betaib*
gammat*mulv*mutb*muti*si + 3*betaib*gammav*mulv*mutb*muti*si) + df(by,t
   )*by*muti*si**2*( - betabi*gammai*
gammat*muti - betabi*gammai*gammav*muti - betabi*gammat*gammav*muti -
   2*betaib*gammab*gammai*mutb*si + 3*
betaib*gammab*gammat*mutb*si + 3*betaib*gammab*gammav*mutb*si + 2*
   betaib*gammab*mulv*mutb*si - 2*betaib*gammai
*gammat*mutb*si - 2*betaib*gammai*gammav*mutb*si + betaib*gammat*mulv*
   mutb*si + betaib*gammav*mulv*mutb*si) +
df(by,t)*iy**4*betaib*mutb**2*(2*gammai**2 - 3*gammai*gammat - 3*gammai
   *gammav - 2*gammai*mulv + gammat*mulv +
 gammav*mulv) + df(by,t)*iy**3*mutb*( - betabi*gammai*gammat*muti -
    betabi*gammai*gammav*muti - betabi*gammat*
gammav*muti + 6*betaib*gammai**2*mutb*si - 9*betaib*gammai*gammat*mutb*
   si - 9*betaib*gammai*gammav*mutb*si - 6
*betaib*gammai*mulv*mutb*si + 3*betaib*gammat*mulv*mutb*si + 3*betaib*
   gammav*mulv*mutb*si + gammab*gammai*
gammat*muti + gammab*gammai*gammav*muti + gammab*gammat*gammav*muti -
```

```
   gammai**2*gammat*muti - gammai**2*gammav
*muti + gammai*gammat*mulv*muti + gammai*gammav*mulv*muti + gammat*
   gammav*mulv*muti) + df(by,t)*iy**2*mutb*si*
( - 2*betabi*gammai*gammat*muti - 2*betabi*gammai*gammav*muti - 2*
   betabi*gammat*gammav*muti + 6*betaib*gammai
**2*mutb*si - 9*betaib*gammai*gammat*mutb*si - 9*betaib*gammai*gammav*
   mutb*si - 6*betaib*gammai*mulv*mutb*si +
 3*betaib*gammat*mulv*mutb*si + 3*betaib*gammav*mulv*mutb*si + 3*gammab
    *gammai*gammat*muti + 3*gammab*gammai*
gammav*muti + 3*gammab*gammat*gammav*muti - 3*gammai**2*gammat*muti -
   3*gammai**2*gammav*muti + 3*gammai*
gammat*mulv*muti + 3*gammai*gammav*mulv*muti + 3*gammat*gammav*mulv*
   muti) + df(by,t)*iy*mutb*si**2*( - betabi*
gammai*gammat*muti - betabi*gammai*gammav*muti - betabi*gammat*gammav*
   muti + 2*betaib*gammai**2*mutb*si - 3*
betaib*gammai*gammat*mutb*si - 3*betaib*gammai*gammav*mutb*si - 2*
   betaib*gammai*mulv*mutb*si + betaib*gammat*
mulv*mutb*si + betaib*gammav*mulv*mutb*si + 3*gammab*gammai*gammat*muti
    + 3*gammab*gammai*gammav*muti + 3*
gammab*gammat*gammav*muti - 3*gammai**2*gammat*muti - 3*gammai**2*
   gammav*muti + 3*gammai*gammat*mulv*muti + 3*
gammai*gammav*mulv*muti + 3*gammat*gammav*mulv*muti) + df(by,t)*mutb*
   muti*si**3*(gammab*gammai*gammat + gammab
*gammai*gammav + gammab*gammat*gammav - gammai**2*gammat - gammai**2*
   gammav + gammai*gammat*mulv + gammai*
gammav*mulv + gammat*gammav*mulv) + df(iy,t,4)*iy**3*mutb**2 + 3*df(iy,
   t,4)*iy**2*mutb**2*si + 3*df(iy,t,4)*iy
*mutb**2*si**2 + df(iy,t,4)*mutb**2*si**3 + df(iy,t,3)*by*iy**3*betaib*
   mutb**2 + 3*df(iy,t,3)*by*iy**2*betaib*
mutb**2*si + 3*df(iy,t,3)*by*iy*betaib*mutb**2*si**2 + df(iy,t,3)*by*
   betaib*mutb**2*si**3 + df(iy,t,3)*iy**3*
mutb**2*(gammai + gammat + gammav + mulv) + 3*df(iy,t,3)*iy**2*mutb**2*
   si*(gammai + gammat + gammav + mulv) +
3*df(iy,t,3)*iy*mutb**2*si**2*(gammai + gammat + gammav + mulv) + df(iy
   ,t,3)*mutb**2*si**3*(gammai + gammat +
gammav + mulv) + df(iy,t,2)*by*iy**3*betaib*mulv*mutb**2 + 3*df(iy,t,2)
   *by*iy**2*betaib*mulv*mutb**2*si + 3*df
(iy,t,2)*by*iy*betaib*mulv*mutb**2*si**2 + df(iy,t,2)*by*betaib*mulv*
```

```
   mutb**2*si**3 + df(iy,t,2)*iy**3*mulv*
mutb**2*(gammai + gammat + gammav) + 3*df(iy,t,2)*iy**2*mulv*mutb**2*si
   *(gammai + gammat + gammav) + 3*df(iy,t
,2)*iy*mulv*mutb**2*si**2*(gammai + gammat + gammav) + df(iy,t,2)*mulv*
   mutb**2*si**3*(gammai + gammat + gammav
) + by**3*iy**4*betaib**3*mutb**2*(gammat + gammav) + by**3*iy**3*
   betaib**2*mutb*(betabi*gammat*muti + betabi*
gammav*muti + 3*betaib*gammat*mutb*si + 3*betaib*gammav*mutb*si -
   gammab*gammat*muti - gammab*gammav*muti) + 3
*by**3*iy**2*betaib**2*mutb*si*(betabi*gammat*muti + betabi*gammav*muti
    + betaib*gammat*mutb*si + betaib*
gammav*mutb*si - gammab*gammat*muti - gammab*gammav*muti) + by**3*iy*
   betaib*si*(betabi**2*gammat*muti**2 +
betabi**2*gammav*muti**2 + 2*betabi*betaib*gammat*mutb*muti*si + 2*
   betabi*betaib*gammav*mutb*muti*si - betabi*
gammab*gammat*muti**2 - betabi*gammab*gammav*muti**2 + betaib**2*gammat
   *mutb**2*si**2 + betaib**2*gammav*mutb
**2*si**2 - 3*betaib*gammab*gammat*mutb*muti*si - 3*betaib*gammab*
   gammav*mutb*muti*si) - by**3*betaib*gammab*
muti*si**2*(betabi*gammat*muti + betabi*gammav*muti + betaib*gammat*
   mutb*si + betaib*gammav*mutb*si) + by**2*
iy**4*betaib**2*mutb**2*(3*gammai*gammat + 3*gammai*gammav - gammat*
   mulv - gammav*mulv) + by**2*iy**3*betaib*
mutb*(2*betabi*gammai*gammat*muti + 2*betabi*gammai*gammav*muti -
   betabi*gammat*mulv*muti - betabi*gammav*mulv
*muti + 9*betaib*gammai*gammat*mutb*si + 9*betaib*gammai*gammav*mutb*si
    - 3*betaib*gammat*mulv*mutb*si - 3*
betaib*gammav*mulv*mutb*si - 2*gammab*gammai*gammat*muti - 2*gammab*
   gammai*gammav*muti + gammab*gammat*mulv*
muti + gammab*gammav*mulv*muti) + by**2*iy**2*betaib*mutb*si*(6*betabi*
   gammai*gammat*muti + 6*betabi*gammai*
gammav*muti + betabi*gammat*gammav*muti - 2*betabi*gammat*mulv*muti -
   2*betabi*gammav*mulv*muti + 9*betaib*
gammai*gammat*mutb*si + 9*betaib*gammai*gammav*mutb*si - 3*betaib*
   gammat*mulv*mutb*si - 3*betaib*gammav*mulv*
mutb*si - 6*gammab*gammai*gammat*muti - 6*gammab*gammai*gammav*muti +
   3*gammab*gammat*mulv*muti + 3*gammab*
gammav*mulv*muti) + by**2*iy*si*(betabi**2*gammai*gammat*muti**2 +
```

```
   betabi**2*gammai*gammav*muti**2 + betabi**2
*gammat*gammav*muti**2 + 4*betabi*betaib*gammai*gammat*mutb*muti*si +
   4*betabi*betaib*gammai*gammav*mutb*muti*
si + betabi*betaib*gammat*gammav*mutb*muti*si - betabi*betaib*gammat*
   mulv*mutb*muti*si - betabi*betaib*gammav*
mulv*mutb*muti*si - betabi*gammab*gammai*gammat*muti**2 - betabi*gammab
   *gammai*gammav*muti**2 - betabi*gammab*
gammat*gammav*muti**2 + 3*betaib**2*gammai*gammat*mutb**2*si**2 + 3*
   betaib**2*gammai*gammav*mutb**2*si**2 -
betaib**2*gammat*mulv*mutb**2*si**2 - betaib**2*gammav*mulv*mutb**2*si
   **2 - 6*betaib*gammab*gammai*gammat*mutb
*muti*si - 6*betaib*gammab*gammai*gammav*mutb*muti*si + 3*betaib*gammab
   *gammat*mulv*mutb*muti*si + 3*betaib*
gammab*gammav*mulv*mutb*muti*si) + by**2*gammab*muti*si**2*( - betabi*
   gammai*gammat*muti - betabi*gammai*
gammav*muti - betabi*gammat*gammav*muti - 2*betaib*gammai*gammat*mutb*
   si - 2*betaib*gammai*gammav*mutb*si +
betaib*gammat*mulv*mutb*si + betaib*gammav*mulv*mutb*si) + by*iy**4*
   betaib*gammai*mutb**2*(3*gammai*gammat + 3
*gammai*gammav - 2*gammat*mulv - 2*gammav*mulv) + by*iy**3*mutb*(betabi
   *gammai**2*gammat*muti + betabi*gammai
**2*gammav*muti - betabi*gammai*gammat*mulv*muti - betabi*gammai*gammav
   *mulv*muti - betabi*gammat*gammav*mulv*
muti + 9*betaib*gammai**2*gammat*mutb*si + 9*betaib*gammai**2*gammav*
   mutb*si - 6*betaib*gammai*gammat*mulv*
mutb*si - 6*betaib*gammai*gammav*mulv*mutb*si - gammab*gammai**2*gammat
   *muti - gammab*gammai**2*gammav*muti +
gammab*gammai*gammat*mulv*muti + gammab*gammai*gammav*mulv*muti +
   gammab*gammat*gammav*mulv*muti) + by*iy**2*
mutb*si*(3*betabi*gammai**2*gammat*muti + 3*betabi*gammai**2*gammav*
   muti + betabi*gammai*gammat*gammav*muti -
2*betabi*gammai*gammat*mulv*muti - 2*betabi*gammai*gammav*mulv*muti -
   2*betabi*gammat*gammav*mulv*muti + 9*
betaib*gammai**2*gammat*mutb*si + 9*betaib*gammai**2*gammav*mutb*si -
   6*betaib*gammai*gammat*mulv*mutb*si - 6*
betaib*gammai*gammav*mulv*mutb*si - 3*gammab*gammai**2*gammat*muti - 3*
   gammab*gammai**2*gammav*muti + 3*gammab
*gammai*gammat*mulv*muti + 3*gammab*gammai*gammav*mulv*muti + 3*gammab*
```

```
   gammat*gammav*mulv*muti) + by*iy*mutb*
si**2*(2*betabi*gammai**2*gammat*muti + 2*betabi*gammai**2*gammav*muti
   + betabi*gammai*gammat*gammav*muti -
betabi*gammai*gammat*mulv*muti - betabi*gammai*gammav*mulv*muti -
   betabi*gammat*gammav*mulv*muti + 3*betaib*
gammai**2*gammat*mutb*si + 3*betaib*gammai**2*gammav*mutb*si - 2*betaib
   *gammai*gammat*mulv*mutb*si - 2*betaib*
gammai*gammav*mulv*mutb*si - 3*gammab*gammai**2*gammat*muti - 3*gammab*
   gammai**2*gammav*muti + 3*gammab*gammai
*gammat*mulv*muti + 3*gammab*gammai*gammav*mulv*muti + 3*gammab*gammat*
   gammav*mulv*muti) + by*gammab*mutb*muti
*si**3*( - gammai**2*gammat - gammai**2*gammav + gammai*gammat*mulv +
   gammai*gammav*mulv + gammat*gammav*mulv)
 + iy**4*gammai**2*mutb**2*(gammai*gammat + gammai*gammav - gammat*mulv
     - gammav*mulv) + 3*iy**3*gammai**2*
mutb**2*si*(gammai*gammat + gammai*gammav - gammat*mulv - gammav*mulv)
   + 3*iy**2*gammai**2*mutb**2*si**2*(
gammai*gammat + gammai*gammav - gammat*mulv - gammav*mulv) + iy*gammai
   **2*mutb**2*si**3*(gammai*gammat +
gammai*gammav - gammat*mulv - gammav*mulv)$

aa_(4) := df(by,t,2)*iy**2*betaib*mutb + df(by,t,2)*iy*betaib*mutb*si +
    2*df(by,t)**2*iy*betaib*muti + 2*df(by
,t)**2*betaib*muti*si - 2*df(by,t)*by*iy**2*betaib**2*mutb + df(by,t)*
   by*iy*betaib*( - 2*betabi*muti - 2*
betaib*mutb*si + 2*gammab*muti + gammat*muti + gammav*muti) + df(by,t)*
   by*betaib*muti*si*(2*gammab + gammat +
gammav) + df(by,t)*iy**2*betaib*mutb*( - 2*gammai + gammat + gammav) +
   df(by,t)*iy*( - 2*betaib*gammai*mutb*si
 + betaib*gammat*mutb*si + betaib*gammav*mutb*si + gammai*gammat*muti +
     gammai*gammav*muti + gammat*gammav*
muti) + df(by,t)*muti*si*(gammai*gammat + gammai*gammav + gammat*gammav
   ) + df(iy,t,3)*iy*mutb + df(iy,t,3)*
mutb*si + df(iy,t,2)*by*iy*betaib*mutb + df(iy,t,2)*by*betaib*mutb*si +
    df(iy,t,2)*iy*mutb*(gammai + gammat +
gammav) + df(iy,t,2)*mutb*si*(gammai + gammat + gammav) - by**2*iy**2*
   betaib**2*mutb*(gammat + gammav) + by**2
*iy*betaib*( - betabi*gammat*muti - betabi*gammav*muti - betaib*gammat*
```

```
   mutb*si - betaib*gammav*mutb*si +
gammab*gammat*muti + gammab*gammav*muti) + by**2*betaib*gammab*muti*si
   *(gammat + gammav) - 2*by*iy**2*betaib*
gammai*mutb*(gammat + gammav) + by*iy*( - betabi*gammai*gammat*muti -
   betabi*gammai*gammav*muti - betabi*
gammat*gammav*muti - 2*betaib*gammai*gammat*mutb*si - 2*betaib*gammai*
   gammav*mutb*si + gammab*gammai*gammat*
muti + gammab*gammai*gammav*muti + gammab*gammat*gammav*muti) + by*
   gammab*muti*si*(gammai*gammat + gammai*
gammav + gammat*gammav) - iy**2*gammai**2*mutb*(gammat + gammav) - iy*
   lx*mulv*mutb*muti*muvt - iy*gammai**2*
mutb*si*(gammat + gammav) - lx*mulv*mutb*muti*muvt*si$

aa_(5) := df(by,t)*by*iy*betaib*muti + df(by,t)*by*betaib*muti*si + df(
   by,t)*iy**2*betaib*mutb + df(by,t)*iy*(
betaib*mutb*si + gammai*muti + gammat*muti) + df(by,t)*muti*si*(gammai
   + gammat) + df(iy,t,2)*iy*mutb + df(iy,
t,2)*mutb*si - by**2*iy**2*betaib**2*mutb + by**2*iy*betaib*( - betabi*
   muti - betaib*mutb*si + gammab*muti) +
by**2*betaib*gammab*muti*si - 2*by*iy**2*betaib*gammai*mutb + by*iy*( -
    betabi*gammai*muti - betabi*gammat*
muti - 2*betaib*gammai*mutb*si + gammab*gammai*muti + gammab*gammat*
   muti) + by*gammab*muti*si*(gammai + gammat
) - iy**2*gammai**2*mutb - iy*vx*mutb*muti*muvt - iy*gammai**2*mutb*si
   - vx*mutb*muti*muvt*si$

aa_(6) := df(by,t)*iy + df(by,t)*si + by*iy*( - betabi + gammab) + by*
   gammab*si - iy*tx*mutb - tx*mutb*si$

aa_(7) :=  - bx + by$

aa_(8) :=  - ax + ay$

aa_(9) :=  - ix + iy$

MODEL ALGEBRAICALLY OBSERVABLE$

PARAMETER VALUES TO SWITCH FROM SYMBOLIC TO NUMERICAL CALCULATIONS$
```

```
b2_ := {mulv=2,
muvt=3,
mutb=5,
muba=7,
muti=11,
gammav=13,
gammat=17,
gammab=19,
gammaa=23,
gammai=29,
betabi=31,
betaib=37,
si=41}$

EXHAUSTIVE SUMMARY (flist)$

flist_ := {mulv - 2,
11*mutb - 5*muti,
muba - 7,
muti*(gammat**2 - 30*gammat + 221),
muti*(betabi - 31),
gammat + gammav - 30,
gammab - 19,
gammaa - 23,
gammai - 29,
betaib - 37,
si - 41}$

MODEL PARAMETER SOLUTION(S)$

g_ := {{betabi=31,
betaib=37,
gammaa=23,
gammab=19,
gammai=29,
gammav=17,
muba=7,
```

```
mulv=2,
mutb=(5*muti)/11,
si=41,
gammat=13},
{betabi=31,
betaib=37,
gammaa=23,
gammab=19,
gammai=29,
gammav=13,
muba=7,
mulv=2,
mutb=(5*muti)/11,
si=41,
gammat=17}}$

MODEL NONIDENTIFIABLE$

IDENTIFIABILITY WITH KNOWN INITIAL CONDITION(S)$

KNOWN INITIAL CONDITION(S)$

ick_ := {lx=1,vx=0,tx=0,bx=0}$

CHARACTERISTIC SET EQUATION(S) EVALUATED AT TIME ZERO AFTER
   SUBSTITUTION OF Y DERIVATIVES AND ICK$

aa_(4) :=  - 13*gammai**2*gammat*si - 169*gammai**2*gammat - 13*gammai
   **2*gammav*si - 169*gammai**2*gammav +
10933*gammai*si + 142129*gammai + 10933*gammat*si + 142129*gammat +
   10933*gammav*si + 142129*gammav - mulv*
muti*muvt*si - 13*mulv*muti*muvt - 316991*si - 4120883$

aa_(5) :=  - gammai**2*si - 13*gammai**2 + 841*si + 10933$

aa_(6) := 0$

aa_(7) := 0$
```

```
aa_(8) :=  - ax + 11$

aa_(9) :=  - ix + 13$

EXHAUSTIVE SUMMARY OF IDENTIFIABILITY WITH KNOWN INITIAL CONDITION(S)$

flisti_ := {mulv - 2,
11*mutb - 5*muti,
muba - 7,
muti*(gammat**2 - 30*gammat + 221),
muti*(betabi - 31),
gammat + gammav - 30,
gammab - 19,
gammaa - 23,
gammai - 29,
betaib - 37,
si - 41,
 - 13*gammai**2*gammat*si - 169*gammai**2*gammat - 13*gammai**2*gammav*
    si - 169*gammai**2*gammav + 10933*
gammai*si + 142129*gammai + 10933*gammat*si + 142129*gammat + 10933*
   gammav*si + 142129*gammav - mulv*muti*muvt
*si - 13*mulv*muti*muvt - 316991*si - 4120883,
 - gammai**2*si - 13*gammai**2 + 841*si + 10933}$

MODEL PARAMETER SOLUTION(S)$

gi_ := {{betabi=31,
betaib=37,
gammaa=23,
gammab=19,
muba=7,
mutb=(5*muti)/11,
muvt=33/muti,
mulv=2,
gammav=17,
si=41,
gammat=13,
```

```
gammai=29},
{betabi=31,
betaib=37,
gammaa=23,
gammab=19,
muba=7,
mutb=(5*muti)/11,
muvt=33/muti,
mulv=2,
gammav=13,
si=41,
gammat=17,
gammai=29}}$

MODEL NONIDENTIFIABLE$
```

3.2

Given the LNP model from equations 2.20a – 2.20f while observing A, and I as state variables, the differential algebra system is several hundred lines long and is not displayed here. It can be accessed in the provided GitHub repository at

`https://github.com/ahadm98/LNPidentifiability`