# Investigation of Inter-Individual Variability in CD8 T Cell Responses with Nonlinear Mixed Effects Models

Chloe Audebert[1,2,3,*], Daphné Laubreton[4], Christophe Arpin[4], Olivier Gandrillon[1,5], Jacqueline Marvel[4,†] , Fabien Crauste[1,6,†]

[1] Inria Dracula, Villeurbanne, France

[2] Sorbonne Université, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions UMR 7598, F-75005 Paris, France.

[3] Sorbonne Université, CNRS, Institut de biologie Paris-Seine (IBPS), Laboratoire de Biologie Computationnelle et Quantitative UMR 7238, F-75005 Paris, France.

[4] Centre International de recherche en Infectiologie, Université de Lyon, INSERM U1111, CNRS UMR 5308, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, 69007 Lyon, France

[5] Laboratory of Biology and Modelling of the Cell, Université de Lyon, ENS de Lyon, Université Claude Bernard, CNRS UMR 5239, INSERM U1210, 46 allée d'Italie, Site Jacques Monod, 69007 Lyon, France

[6] Univ. Bordeaux, CNRS, Bordeaux INP, IMB, UMR 5251, F-33400, Talence, France

**Abstract**

To develop vaccines it is mandatory yet challenging to account for inter-individual variability during immune responses. Even in laboratory mice, T cell responses of single individuals exhibit a high heterogeneity that may come from genetic backgrounds, intra-specific processes (*e.g.* antigen-processing and presentation) and immunization protocols.

We propose to account for inter-individual variability in CD8 T cell responses in mice by using a dynamical model and a statistical, nonlinear mixed effects model. Average and

*corresponding author: chloe.audebert@sorbonne-universite.fr

†co-last authors

individual dynamics during a CD8 T cell response are characterized in different immunization contexts (vaccinia virus and tumor). We identify biological processes more likely to be affected by the immunization and to generate inter-individual variability. The robustness of the model is assessed by confrontation to new experimental data, and it proves able to predict tumor volume dynamics as well as individual dynamics.

Our approach allows to investigate immune responses in various immunization contexts, when measurements are scarce or missing, and contributes to a better understanding of variability in CD8 T cell immune responses. In particular, the prediction of tumor volume dynamics based solely on information on CD8 T cell counts may lead to tumor therapy improvement.

*Keywords:* T cell response ; Inter-individual variability ; ODE dynamical model ; Nonlinear mixed effects models ; Prediction of tumor evolution.

# 1 Introduction

The immune response is recognized as a robust system able to counteract invasion by diverse pathogens (Fischer and Raussel, 2016; Wong and Germain, 2018). However, as a complex biological process the dynamical behavior of its cellular components exhibits a high degree of variability affecting their differentiation, proliferation or death processes. Indeed, the frequency of antigen-specific T cells and their location relative to pathogen invasion will affect the dynamic of the response (Estcourt et al., 2005; Wong and Germain, 2018; Xiao et al., 2007). Similarly, the pathogen load and virulence as well as the host innate response will affect the T cell response (Iwasaki and Medzhitov, 2015). Finally, at the cellular level, variation in protein content can also generate heterogeneous responses (Feinerman et al., 2008). Genetic variability of the numerous genes controlling the immune response will also be a source of variability among individuals (Fischer and Raussel, 2016). However, even among genetically identical individuals, the response to the same infection can result in highly heterogeneous dynamics (Althaus et al., 2007; Grau et al., 2018; Murali-Krishna et al., 1998).

Cytotoxic CD8 T cells play an essential role in the fight against pathogens or tumors as they are able to recognize and eliminate infected or transformed cells. Indeed, following encounter, in lymphoid organs, of antigen-presenting cells loaded with pathogen or tumor

derived antigens, quiescent naive CD8 T cells will be activated. This leads to their proliferation and differentiation in effector cells that have acquired the capacity to kill infected or tumoral cells, and to their ultimate differentiation in memory cells (Crauste et al., 2017; Youngblood et al., 2017). The CD8 T cell immune response is a robust process that, once engaged by antigenic stimulations, will proceed up to the memory phase, even if the antigenic stimulation is shortened. Hence, the auto-pilot denomination given to this process (Kaech and Ahmed, 2001; Stipdonk et al., 2001). However, it is highly variable as illustrated by experimental measurements of cell counts: dynamics of the responses (timing, cell counts) may differ from one individual to another (Miller et al., 2008; Precopio et al., 2007; Xiao et al., 2007), but also depending on the immunogen (Althaus et al., 2007; Estcourt et al., 2005; Murali-Krishna et al., 1998).

The role of genome variability in explaining inter-individual variations of T cell responses has been recently investigated (Ferraro et al., 2014; Li et al., 2016) but provided limited understanding of the observed heterogeneity. Li et al. (2016) focused on correlations between gene expression and cytokine production in humans, and identified a locus associated with the production of IL-6 in different pathogenic contexts (bacteria and fungi). Ferraro et al. (2014) investigated inter-individual variations based on genotypic analyses of human donors (in healthy and diabetic conditions) and identified genes that correlate with regulatory T cell responses.

To our knowledge, inter-individual variability characterized by heterogeneous cell counts has been mostly ignored in immunology, put aside by focusing on average behaviors of populations of genetically similar individuals. The use of such methodology, however, assumes that variability is negligible among genetically similar individuals, which is not true (Althaus et al., 2007; Badovinac et al., 2007; Crauste et al., 2017).

In this work, we propose to study inter-individual variability based on CD8 T cell counts with nonlinear mixed effects models (Delyon et al., 1999; Kuhn and Lavielle, 2005; Lavielle, 2014). In these models, instead of considering a unique set of parameter values as characteristic of the studied data set, a population approach is used based on distributions of parameter values. All individuals are assumed to be part of the same population, and as so they share a common behavior (average behavior) while they keep individual behaviors (random effect).

Nonlinear mixed effects models are well adapted to repeated longitudinal data. They aim at characterizing and understanding "typical behaviors", and as a consequence inter-individual variations. T cell count measurements, obtained over the course of a response (few weeks), and the large variability observed in them represent a case study for this approach.

Nonlinear mixed effects models have been used to analyze data in various fields (Davidian and Giltinan, 2003), especially in pharmakokinetic studies, and more recently to model cell to cell variability (Almquist et al., 2015; Llamosi et al., 2016) or to study tumor growth (Benzekry et al., 2014; Ferenci et al., 2017). In immunology, Keersmaekers et al. (2018) have recently studied the differences between two vaccines with nonlinear mixed effects models and ordinary differential equation (ODE) models for T and B cells. Jarne et al. (2017) and Villain et al. (2018) have investigated the effect of IL7 injections on HIV+ patients to stimulate the CD4 T cell response using also nonlinear mixed effects models and ODEs, and identified biological processes accounting for inter-individual variability.

A number of models of the CD8 T cell response based on ODEs have been proposed over the years. Of particular relevance here is the work of De Boer et al. (2001), where the model accounts for activated and memory cells but the influence of the immunogen is imposed. Antia et al. (2003) proposed a model based on partial differential equations, that includes immunogen effects and dynamics of naive, effector and memory cells. These works describe different subpopulations of CD8 T cells, however most of the time only total CD8 T cell counts are available to validate the models. In Crauste et al. (2017), the authors generated cell counts for four subpopulations of CD8 T cells in mice that they used to identify the most likely differentiation pathway of CD8 T cells after immunogen presentation. This approach has led to a model of the average CD8 T cell dynamics in mice after immunization and its representation as a set of nonlinear ODEs. The model consists in a system of ODEs describing the dynamics of naive, early effector, late effector, and memory CD8 T cell subsets and the immunogen.

The goal of this article is to explore the ability of a mathematical model to describe the inter-individual variability observed in CD8 T cell responses in different immunization contexts, by considering distributions of parameters within the population (nonlinear mixed effects model). We will first select a model of the CD8 T cell immune response dynamics accounting

for variability in cell counts by using synthetic then experimental data generated in different immunization contexts. Second we will establish that the immunogen-dependent heterogeneity influences the early phase of the response (priming, activation of naive cells, cellular expansion). Finally, we will show that besides its ability to reproduce CD8 T cell response dynamics our model accounts for relevant immunogen dynamics (tumor volume evolution) and is able to predict dynamics of responses to similar immunizations, hence providing an efficient tool for investigating CD8 T cell dynamics and inter-individual variability.

# 2 Material, Methods and Models

## 2.1 Data

**Experimental Models.** C57BL/6 mice (C57BL6/J) and CD45.1+ C57BL/6 mice (B6.SJL-Ptprc$^{a}$Pepc$^{b}$/BoyCrl) were purchased from CRL. F5 TCR-tg mice recognizing the NP68 epitope were crossed to a CD45.1+ C57BL/6 background (B6.SJL-Ptprc$^{a}$Pepc$^{b}$/BoyCrl-Tg(CD2-TcraF5,CD2-TcrbF5)1Kio/Jmar) (Jubin et al., 2012). They have been crossed at least 13 times on the C57BL6/J background. All mice were homozygous adult 6-8-week-old at the beginning of experiments. They were healthy and housed in our institute's animal facility under Specific Pathogen-Free conditions.

Age- and sex-matched litter mates or provider's delivery groups, which were naive of any experimental manipulation, were randomly assigned to 4 experimental groups (of 5 mice each) and co-housed at least for one week prior to experimentation. Animals were maintained in ventilated enriched cages at constant temperature and hygrometry with 13hr/11hr light/dark cycles and constant access to 21 kGy-irradiated food and acid (pH $= 3 \pm 0.5$) water. All experimental procedures were approved by an animal experimentation ethics committee (CECCAPP; Lyon, France), and accreditations have been obtained from French government.

**Vaccinia Virus (VV) Immunization.** $2 \times 10^5$ naive CD8 T cells from CD45.1+ F5 mice were transferred by retro-orbital injection in, 6-8-week-old congenic CD45.2+ C57BL/6 mice briefly anaesthetized with 3% isoflurane. The day after deeply Xylazin/ Ketamin-anaesthetized recipient mice were inoculated intra-nasally with $2 \times 10^5$ pfu of a vaccinia virus expressing the

NP68 epitope (VV-NP68) provided by Pr. A.J. McMichael (Jubin et al., 2012).

**Tumor Immunization.** $2 \times 10^5$ naive CD8 T cells from CD45.1+ F5 mice were transferred by retro-orbital injection in 6-8-week-old congenic CD45.2+ C57BL/6 mice briefly anaesthetized with 3% isoflurane. The day after, recipients were subcutaneously inoculated with $2.5 \times 10^6$ EL4 lymphoma cells expressing the NP68 epitope (EL4-NP68) provided by Dr. T.N.M. Schumacher (de Brito et al., 2011).

**Phenotypic Analyses.** Mice were bled at intervals of at least 7 days. Blood cell suspensions were cleared of erythrocytes by incubation in ACK lysis solution (TFS). Cells were then incubated with efluor780-coupled Fixable Viability Dye (eBioscience) to label dead cells. All surface stainings were then performed for 45 minutes at 4°C in PBS (TFS) supplemented with 1% FBS (BioWest) and 0.09% NaN3 (Sigma-Aldrich). Cells were fixed and permeabilized with the Foxp3-fixation and permeabilization kit (eBioscience) before intra-cellular staining for one hour to overnight. The following mAbs(clones) were utilized: Bcl2(BCL/10C4), CD45.1(A20) and CD45(30-F11) from Biolegend, Mki67(SolA15), CD27(LG.7F9) and CD8(53.6.7) from eBioscience, and CD44 (IM7.8.1) from Miltenyi. Samples were acquired on a FACS LSR Fortessa (BD biosciences) and analyzed with FlowJo software (TreeStar).

**CD8 T Cell Differentiation Stages.** For both immunizations (VV and Tumor), phenotypic cell subsets based on Mki67-Bcl2 characterization (Crauste et al., 2017) have been identified and the corresponding cell counts measured in blood, from day 4 post-inoculation (pi) up to day 28pi, 32pi, 46pi, or 47pi depending on the experiment (VV and Tumor data sets 1, Table 1). Naive cells are defined as CD44-Mki67-Bcl2+ cells, early effector cells as CD44+Mki67+Bcl2- cells, late effector cells as CD44+Mki67-Bcl2- cells, and memory cells as CD44+Mki67-Bcl2+ cells.

**Tumor volumes.** Twenty mice inoculated with EL4 lymphoma cells had tumor volumes measured daily between day 4pi to 13pi (Tumor data set 3, Table 1). Tumors were subcutaneously inoculated on the back of the animals allowing tumor volume measurements. Tumor

| Short Name | Description |
|---|---|
| VV data set 1 | CD8 T cell counts of 59 individual mice inoculated intranasally with $2 \times 10^5$ pfu of a vaccinia virus expressing the NP68 epitope ; naive, early and late effector, and memory cell counts have been measured up to day 47pi |
| VV data set 2 | Similar to VV data set 1 (15 individual mice) ; CD8 T cell counts of naive, early and late effector, and memory cells have been measured following VV immunization, up to day 42pi |
| Tumor data set 1 | CD8 T cell counts of 55 individual mice subcutaneously inoculated with $2.5 \times 10^6$ EL4 lymphoma cells expressing the NP68 epitope ; naive, early and late effector, and memory cell counts have been measured up to day 47pi |
| Tumor data set 2 | Similar to Tumor data set 1 (20 individual mice); CD8 T cell counts of naive, early and late effector, and memory cells have been measured following Tumor immunization, up to day 42pi |
| Tumor data set 3 | Tumor volumes measured daily between day 4pi to 13pi for 20 individual mice of the Tumor data set 1 |
| Synth data set 1 | Synthetic data set generated with System (1), consisting in CD8 T cell counts of naive, early and late effector, and memory cells on days 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30pi for 10 individuals |
| Synth data set 2 | Similar to Synth data set 1, except that System (2) is used to generate the data |

Table 1: Data sets (details in the text of sections 2.1, 2.4 and 2.5).

length ($l$), width ($w$) and thickness ($t$) were measured and the tumor volume was estimated with the formula $\pi lwt/6$, assuming tumors are ellipsoids (Tomayka and Reynolds, 1989).

## 2.2 Nonlinear mixed effects models

Nonlinear mixed effects models allow a description of inter-individual heterogeneity within a population of individuals (here, mice). The main idea of the method is to consider that since all individuals belong to the same population they share common characteristics. These common characteristics are called "fixed effects" and characterize an average behavior of the population. However, each individual is unique and thus differs from the average behavior by a specific value called "random effect".

This section briefly describes our main hypotheses. Details on the method can be found in Delyon et al. (1999), Kuhn and Lavielle (2005), Samson and Donnet (2007), Lavielle (2014).

Each data set $\{y_{i,j},\ i = 1, ..., N_{ind},\ j = 1, ..., n_i\}$ is described as follows,

$$y_{i,j} = f(x_{i,j}, \psi_i) + a\varepsilon_{i,j},$$

where $y_{i,j}$ is the $j^{th}$ observation of individual $i$, $N_{ind}$ is the number of individuals within the population and $n_i$ is the number of observations for the $i^{th}$ individual.

The function $f$ accounts for individual dynamics generated by a mathematical model, in this work $f$ is associated with the solution of a system of ordinary differential equations (ODE), see section 2.4. The function $f$ depends on known variables, denoted by $x_{i,j}$, and parameters of the $i^{th}$ individual, denoted by $\psi_i$. Here, known variables are CD8 T cell subpopulations and time.

Individual parameters $\psi_i$ are assumed to be split into fixed effects (population-dependent effects, average behavior) and random effects (individual-dependent effects). Denote $\psi_i^k$ the $k$-th parameter characterizing individual $i$. Then

$$\log(\psi_i^k) = \log(p_{pop}^k) + \eta_i^k,$$

where vector $p_{pop} = (p_{pop}^k)_k$ models the average behavior of the population, and $\eta_i = (\eta_i^k)_k$ represents how the individual $i$ differs from this average behavior. Variables $\eta_i^k \sim \mathcal{N}(0, \omega_k^2)$, and they are assumed independent and identically distributed. The variance $\omega_k^2$ quantifies the

variability of the $k$-th parameter within the population. From now we will denote by $\omega^2$ the vector of variances $(\omega_k^2)_k$. Parameters $\psi_i$ are assumed to follow a log-normal distribution to ensure their positivity.

The residual errors, combining model approximations and measurement noise, are denoted by $a\varepsilon_{i,j}$. They quantify how the model prediction is close to the observation. Residual errors are assumed independent, identically and normally distributed, *i.e* $\varepsilon_{i,j} \sim \mathcal{N}(0,1)$. Moreover, the random effects $\eta_i$ and the residual errors $a\varepsilon_{i,j}$ are mutually independent. In this work, we assume a *constant* error model, with a constant $a$. This error parameter is estimated for each subpopulation (naive cells - $a_N$ ; early effector cells - $a_E$ ; late effector cells - $a_L$ ; memory cells - $a_M$).

In what follows, we will write that a parameter is *fixed within the population* if all individuals have the same value for this parameter. On the contrary, if the variance $\omega_k^2$ of a parameter is non-zero, then this parameter will account for inter-individual variability within the population.

### 2.3 Parameter Estimation

Parameter values are estimated with Stochastic Approximation Expectation-Maximization (SAEM) algorithm. The SAEM algorithm is available in Monolix (2018).

**Population and individual parameters.** Under the previous assumptions (section 2.2), cell population dynamics (average behavior and inter-individual variability) are described by parameters: $p_{pop}$, $\omega^2$ and $a$. These parameters are estimated by maximizing the likelihood with the SAEM algorithm.

Once these parameters have been estimated, each individual vector of parameters $\psi_i$ is estimated by maximizing the conditional probabilities $\mathbb{P}(\psi_i|y_{i,j};\hat{p}_{pop},\hat{\omega}^2,\hat{a})$, where $\hat{x}$ denotes the estimated value of $x$.

Both estimations are performed with Monolix software (Monolix, 2018). Files to run the algorithm (including all algorithm parameters) are available in Supplementary File 3.

**Covariates.** In section 3.3, we will study whether differences observed in parameter values between VV and Tumor data sets (Table 1) are only related to random sampling, or if they can be explained by the immunogen. If so, then this should be reflected in the values of some parameters.

To tackle this question, we first pool together VV and Tumor data sets 1. Second, using this full data set, we estimate parameter values by assuming that fixed effects of some Tumor-associated parameters are different from those of the corresponding VV-associated parameters. If these differences in parameter values are significantly different from zero, then they are caused by the immunogen.

This question is translated into the mixed effects model by introducing categorical covariates. It assumes that a given individual parameter vector $\psi_i$ follows a probability distribution with a different mean depending if the individual is in Tumor or VV data set. We write

$$\log(\psi_i^k) = \log(p_{pop}^k) + \beta^k c_i + \eta_i^k,$$

where $c_i$ equals 0 if individual $i$ is in VV data set 1 and 1 for individuals in Tumor data set 1, and $\beta = (\beta^k)_k$ is a vector of covariate parameters (it quantifies the difference between values of the mean parameter of both data sets). We test whether the estimated covariate parameter $\hat{\beta}$ is significantly different from zero with a Wald test, using Monolix (2018) software, and we use a $p$-value threshold at 0.05.

Parameters $(p_{pop}, \omega^2, a, \beta)$ are then characterizing cell population dynamics for both VV and Tumor immunogens. If the estimated vector $\hat{\beta}$ is significantly different from zero, then part of the experimentally observed variability could be explained by the immunogen.

## 2.4 Model selection

**ODE model of CD8 T cells dynamics.** We introduce the system of ODE that accounts for each individual behavior. We use the model in Crauste et al. (2017), that describes CD8 T cell subpopulation dynamics as well as the immunogen load dynamics in primary immune

responses, as follows

$$\begin{cases} \dot{N} &= -\mu_N N - \delta_{NE} I N, \\ \dot{E} &= \delta_{NE} I N + \rho_E I E - \left[\mu_E E + \delta_{EL}\right] E, \\ \dot{L} &= \delta_{EL} E - \left[\mu_L^L L + \mu_L^E E + \delta_{LM}\right] L, \\ \dot{M} &= \delta_{LM} L, \\ \dot{I} &= \left[\rho_I I - \mu_I^E E - \mu_I^L L - \mu_I\right] I. \end{cases} \tag{1}$$

The variables $N$, $E$, $L$ and $M$ denote the four CD8 T cell subpopulation counts, naive, early effector, late effector, and memory cells respectively (see section 2.1), and $I$ is the immunogen load.

The immunogen load dynamics are normalized with respect to the initial amount (Crauste et al., 2015, 2017), so $I(0) = 1$. The initial amounts of CD8 T cell counts are $N(0) = 10^4$ cells, $E(0) = 0$, $L(0) = 0$ and $M(0) = 0$.

Parameters $\delta_k$ are the differentiation rates, with $k = NE$, $EL$ or $LM$ for differentiation from naive to early effector cells, from early effector to late effector cells and from late effector to memory cells, respectively.

Death parameters are denoted by $\mu_k$, where $k = N$, $E$ and $I$ for the death of naive, early effector cells and the immunogen respectively. Notations $\mu_X^Y$ for some mortality-related parameters refer to parameters $\mu_{XY}$ in Crauste et al. (2017): the index $X$ refers to the CD8 T cell population or the immunogen that dies, and the exponent $Y$ to the CD8 T cell population responsible for inducing death.

Proliferation parameters of early effector cells and the immunogen are respectively denoted by $\rho_E$ and $\rho_I$.

System (1) has been introduced and validated in Crauste et al. (2017). This system will be simplified here to obtain locally identifiable parameters on ideal synthetic data.

**Synthetic data.** Using System (1), we generate a set of data associated to solutions of the model, where all the parameters are drawn from known log-normal distributions. These data consist of time points and CD8 T cell counts for the 4 subpopulations and the immunogen

load. These are called *synthetic data*, and this set of data is referred to as Synth data set 1 (Table 1).

In this first step, each model parameter is assumed to account for inter-individual variability, so no parameter is fixed within the population, and each parameter $p_k$ satisfies $\log(p_k) \sim \mathcal{N}(\log(m_k), 0.2)$. The standard deviation is fixed to the value 0.2 to generate heterogeneity, and values of medians $m_k$ are given in Table S1. In a second step, after reducing the number of parameters (see section 3.1), we generate another set of synthetic data (Synth2 data set, Table 1), with the same assumptions and methods.

We generate synthetic data for 10 individuals, whose cell counts are sampled at days 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30pi (cf. Figure S1). In agreement with real biological data, we assume that all cell counts below 100 cells are not measured, and account for missing data. For the immunogen load, values lower than 0.1 are also not considered.

**Model selection on biological data.** Using the system selected on the Synth data set 2 (Table 1) and experimental data presented in section 2.1 (VV data set 1, then Tumor data set 1, see Table 1), we perform a parameter estimation (see section 2.3).

## 2.5 A posteriori model validation on biological data

In sections 3.4 and 3.5, the model selected on real biological data is compared to data that were not used for parameter estimation. These data are presented hereafter.

**Tumor volume.** Tumor volume was measured for one experiment (Tumor data set 3, Table 1), as described in section 2.1. Tumors grown from EL4 lymphoma cells are known to develop on one site and show almost no metastases (Boissonnas et al., 2004; Vetvicka et al., 2009). Consequently, we assume that immunogen load dynamics and the tumor volume evolution are related. Therefore we can compare experimentally measured tumor volumes with immunogen load dynamics generated by the model.

Since the immunogen load dynamics remain unknown in experimental data, the initial immunogen count has been normalized in the model (Crauste et al., 2015, 2017). Consequently, we can only qualitatively compare immunogen load dynamics predicted by the model with

tumor volume measurements. Features of interest are then the time when the immunogen load/tumor volume reaches its maximum, as well as the elimination rate of the immunogen load/tumor. To be able to compare these features of interest, tumor volumes and simulated immunogen load values are normalized by their maximum value.

**Additional experiment.** In order to assess the model ability to characterize and predict immune response dynamics we compare our results to additional experiments, VV data set 2 and Tumor data set 2 (see Table 1 and section 2.1), similar to the ones used to estimate parameters (VV and Tumor data sets 1). CD8 T cell counts of naive, early and late effector, and memory cells have been measured following VV and tumor immunizations, on days 4, 6, 7, 8, 11, 13, 15, 21, 28, 42pi.

The probability distribution of parameters (population-dependent, fixed effects) are known since we have estimated them on VV and Tumor data sets 1 (section 2.4). Therefore we use them to estimate the individual parameter values that fit individual behaviors of these new data sets (see section 2.3).

# 3 Results

## 3.1 Model selection on synthetic data

Using Synth data set 1 (ideal data generated by System (1), see section 2.4), System (1) is reduced in order for population parameters to be locally identifiable. Parameter estimation (see section 2.3) is performed with SAEM algorithm (Monolix, 2018) on the synthetic data, and it leads to a reduction of the initial 12-parameters System (1) to the 9-parameters System (2), as explained herefater.

One way to obtain parameter identifiability is to assume known 3 parameters: the additional death rates of late effector cells ($\mu_L^E$) and of the immunogen ($\mu_I$, $\mu_I^E$). First their initial true median values are used, see Table S1. Second, we observe that setting these parameters to zero does not impair neither the quality of fit nor the parameter estimation (not shown), so they are removed from the model. Since only 9 parameters are required (see Table S1) to

reproduce the expected individual dynamics, we obtain System (2),

$$\begin{cases} \dot{N} &= -\mu_N N - \delta_{NE} I N, \\ \dot{E} &= \delta_{NE} I N + \rho_E I E - [\mu_E E + \delta_{EL}] E, \\ \dot{L} &= \delta_{EL} E - [\mu_L L + \delta_{LM}] L, \\ \dot{M} &= \delta_{LM} L, \\ \dot{I} &= [\rho_I I - \mu_I L] I. \end{cases} \tag{2}$$

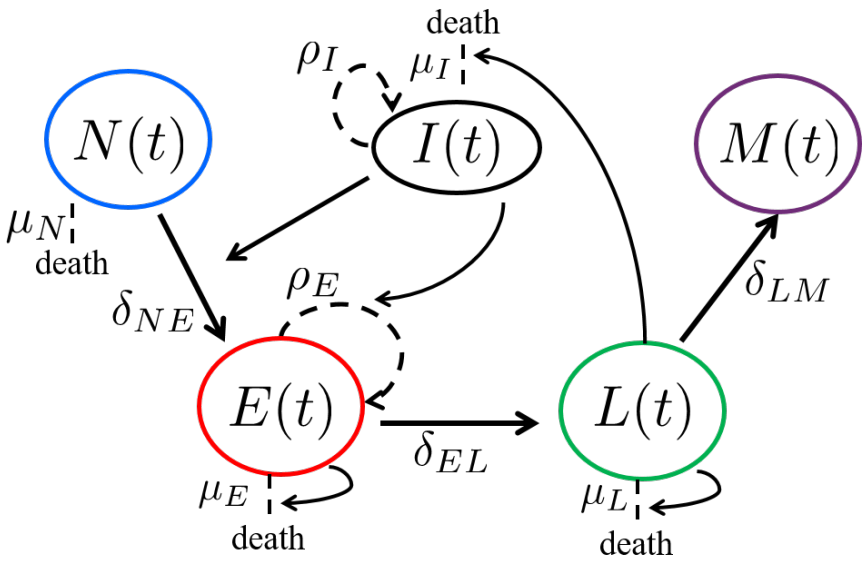
For the sake of simplicity the parameters are renamed in System (2): $\mu_L^L = \mu_L$ and $\mu_I^L = \mu_I$. Figure 1.A displays a schematic representation of System (2).

Additionally, among the 9 parameters characterizing System (2), we wondered whether all parameters had to vary within the population to explain the observed dynamics. Indeed, we want to avoid redundant sources of variability. To our knowledge it is impossible to determine whether differentiation of CD8 T cells from one subpopulation to another one varies between individuals. Moreover, the autopilot process discovered by Kaech and Ahmed (2001) and Stipdonk et al. (2001) indicates that differentiation of naive in memory CD8 T cells is a robust process. Therefore, we assume that the differentiation rates (parameters $\delta_{NE}$, $\delta_{EL}$ and $\delta_{LM}$) are fixed within the population.

The parameter estimation is then performed on Synth data set 2 (Table 1) for System (2), with 3 parameters fixed within the population over a total of 9 parameters. Results are only slightly impacted (Figure S1, black *vs* red curves), most individual behaviors are similarly reproduced. Some measurements are not well captured but the overall dynamics is consistent with the population behavior since population parameters are similar for both estimations (see Table S2). One may note that estimated residual errors are increased (Table S2) but remain very small, attesting that this new assumption is reasonable.

From this study, we select the model described by System (2) with 3 parameters ($\delta_{NE}$, $\delta_{EL}$ and $\delta_{LM}$) fixed within the population and 6 parameters varying between individuals ($\mu_N$, $\mu_E$, $\mu_L$, $\mu_I$, $\rho_E$, $\rho_I$) (Figure 1.A).

**(A) Schematic representation of System** (2)




**(B) Schematic representation of System** (3)

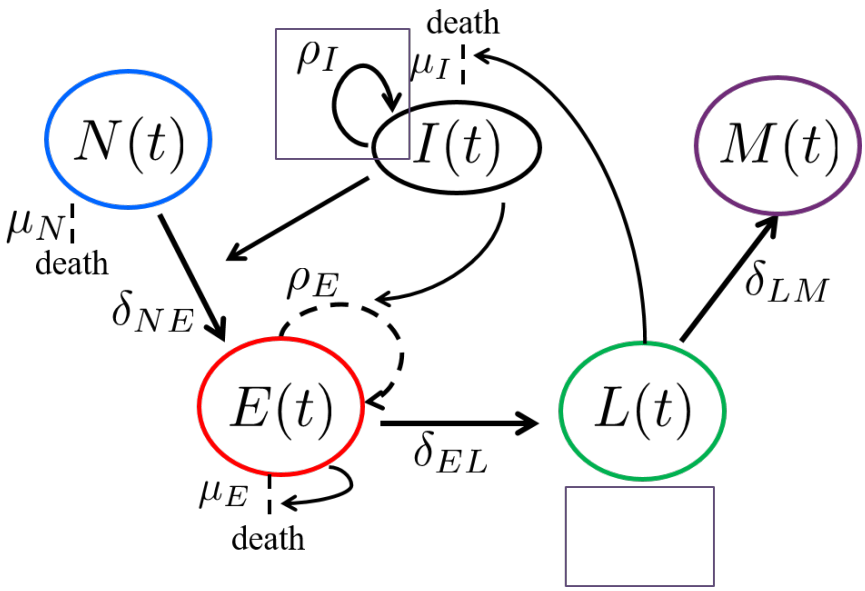



Figure 1: Schematic CD8 T cell differentiation diagram following immunization. (A) Schematic representation of System (2). (B) Schematic representation of System (3), reduced based on *in vivo* data. Rectangles highlight the differences between the two models. In both cases, dashed lines represent individual-dependent parameters while solid lines correspond to parameters fixed within the population.

### 3.2 A model of CD8 T cell dynamics accounting for *in vivo* inter-individual heterogeneity

Parameter estimation for System (2) is performed as detailed in section 2.4 using VV data set 1 (see section 2.1). Using *in vivo* data to estimate parameter values provides a priori less information than synthetic data. Hence, it may be necessary to reduce the number of parameters, either fixed or varying, to be estimated, similarly to what has been done in the previous section.

Contrary to the estimation performed on synthetic data in section 3.1, using real biological measurements from VV data set 1 leads to an estimated value of parameter $\mu_L$ close to zero ($10^{-10}$), see Table 2, step 1. The estimation is therefore performed again with the assumption $\mu_L = 0$. We observe that the other parameters are not impacted (Table 2, step 2). Parameter $\mu_L$ is hence set to zero in System (2). Setting this parameter to zero does not mean that late effector cells do not die *in vivo*, but that the available data are not rich enough to estimate this parameter value. In other words, with the information contained into the data set, this parameter is not required to describe the CD8 T cell immune response.

Despite a non-negligible estimated standard deviation of parameter $\rho_I$ (Table 2, step 2), the empirical standard deviation of the individual parameters $\rho_I$ is low ($2\times10^{-4}$). Consequently we assume that parameter $\rho_I$ is fixed within the population. The estimation with this assumption does not impact other parameter values (Table 2, step 3). Moreover, the estimated residual errors (Table 2) are the same at each step showing that the successive simplifications do not impact the estimated error of fit between observations and the system solutions.

From these estimations, we obtain System (3) that enables to describe VV data set 1 and the inter-individual variability it displays,

$$\begin{cases} \dot{N} &= -\mu_N N - \overline{\delta}_{NE} I N, \\ \dot{E} &= \overline{\delta}_{NE} I N + \rho_E I E - \left[\mu_E E + \overline{\delta}_{EL}\right] E, \\ \dot{L} &= \overline{\delta}_{EL} E - \overline{\delta}_{LM} L, \\ \dot{M} &= \overline{\delta}_{LM} L, \\ \dot{I} &= \left[\overline{\rho}_I I - \mu_I L\right] I, \end{cases} \tag{3}$$

| Param. | Units | Step 1<br>System (2)<br>(9 param.)<br>(6 v.p.) | Step 2<br>$+ \mu_L = 0$<br>(8 param.)<br>(5 v.p.) | Step 3<br>$+ \rho_I$ fixed<br>(8 param.)<br>(4 v.p.) |
|---|---|---|---|---|
| *Population parameters* | | | | |
| $\mu_N$ | day$^{-1}$ | 0.66 | 0.66 | 0.66 |
| $\omega_{\mu_N}$ | day$^{-1}$ | 0.3 | 0.3 | 0.3 |
| $\delta_{NE}$ | day$^{-1}$ | 0.0004 | 0.0004 | 0.0007 |
| $\rho_E$ | day$^{-1}$ | 1.1 | 1.1 | 1.0 |
| $\omega_{\rho_E}$ | day$^{-1}$ | 0.2 | 0.2 | 0.2 |
| $\mu_E$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 6.4 | 6.2 | 6.0 |
| $\omega_{\mu_E}$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 0.8 | 0.8 | 0.9 |
| $\delta_{EL}$ | day$^{-1}$ | 0.1 | 0.1 | 0.1 |
| $\mu_L$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 9.4 $10^{-5}$ | - | - |
| $\omega_{\mu_L}$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 3.1 | - | - |
| $\delta_{LM}$ | day$^{-1}$ | 0.09 | 0.09 | 0.09 |
| $\rho_I$ | day$^{-1}$ | 0.09 | 0.08 | 0.1 |
| $\omega_{\rho_I}$ | day$^{-1}$ | 0.03 | 0.03 | - |
| $\mu_I$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 2.0 | 1.9 | 2.4 |
| $\omega_{\mu_I}$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 1.1 | 1.2 | 1.0 |
| *Residual errors* | | | | |
| $a_N$ | cell counts (log10) | 0.2 | 0.2 | 0.2 |
| $a_E$ | cell counts (log10) | 0.4 | 0.4 | 0.4 |
| $a_L$ | cell counts (log10) | 0.5 | 0.5 | 0.5 |
| $a_M$ | cell counts (log10) | 0.3 | 0.3 | 0.3 |

Table 2: Steps in estimating parameter values using VV data set 1. Step 1 (third column) shows values estimated using System (2). Step 2 (fourth column) shows values estimated when $\mu_L$ (grey) is set to zero in System (2). Step 3 (fifth column) shows values estimated with the final modification of System (2), that is System (3), where the parameter $\rho_I$ is now fixed within the population (grey). Columns 3 to 5, 'param.' stands for parameters, and 'v.p.' stands for varying parameter within population.

with a bar highlighting fixed parameters within the population: parameters $\mu_N$, $\mu_E$, $\mu_I$ and $\rho_E$ are varying within the population, whereas $\overline{\rho}_I$, $\overline{\delta}_{NE}$, $\overline{\delta}_{EL}$ and $\overline{\delta}_{LM}$ are fixed for all individuals (Figure 1.B).

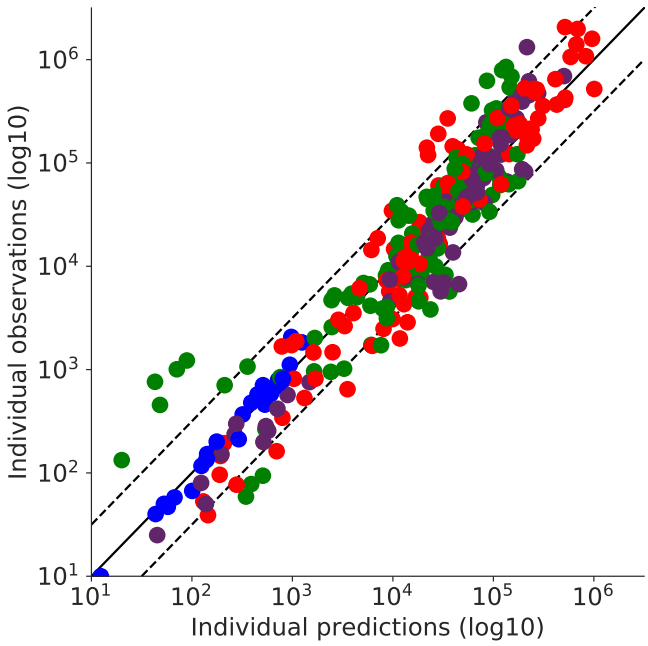
Figure 2.A shows the good agreement between model predictions and individual measurements for each CD8 T cell subpopulation. Model predictions are the results of numerical simulations of System (3) performed for each individual with the estimated individual parameter values. Except for few underestimated measurements of late effector cell counts, all cell counts are well matched with the model. Parameter values are given in Table 3.

The five underestimated measurements of late effector cell counts correspond to day 4pi measurements. They are the earliest late effector cell counts that were measured, and our model underestimates them by a 10-fold factor approximately. It is noticeable that at day 4pi blood is not a very good mirror of CD8 T cell dynamics since most cells are still located in lymphoid organs. Hence experimental values must be considered with caution. Additionally, late effector cell counts on day 4pi were not available in Crauste et al. (2017) to validate the dynamical model, and that may also explain the underestimation. However, it does not impact the overall estimation of CD8 T cell dynamics, especially at later time points (see Figure 2.A).

Figure 3 shows the estimated dynamics of early- and late-effector and memory cells of two individuals. One individual (dashed curves) was monitored on day 7pi, 15pi and 47pi leading to three measurements points for late effetor cells and two for early effector and memory cells. Estimated dynamics are in agreement with what is expected, especially regarding the time and height of the peak of the response and the following contraction phase. The other individual (solid curves) had cell count measurements only on day 8pi, yet the estimated dynamics correspond to an expected behavior, which could not have been obtained by fitting this individual alone. Hence we are able to simulate likely dynamics even with a small amount of data points, thanks to the use of nonlinear mixed effects models. By focusing first on the population dynamics (based on a collection of individual dynamics), the method enables to recover the entire individual dynamics. This is a huge advantage when data sampling frequency is low.

Similar good results are obtained for Tumor data set (see Figure 2.B). Therefore System (3) enables to describe inter-individual variability in different immunization contexts, here VV

**(A) VV**




**(B) Tumor**

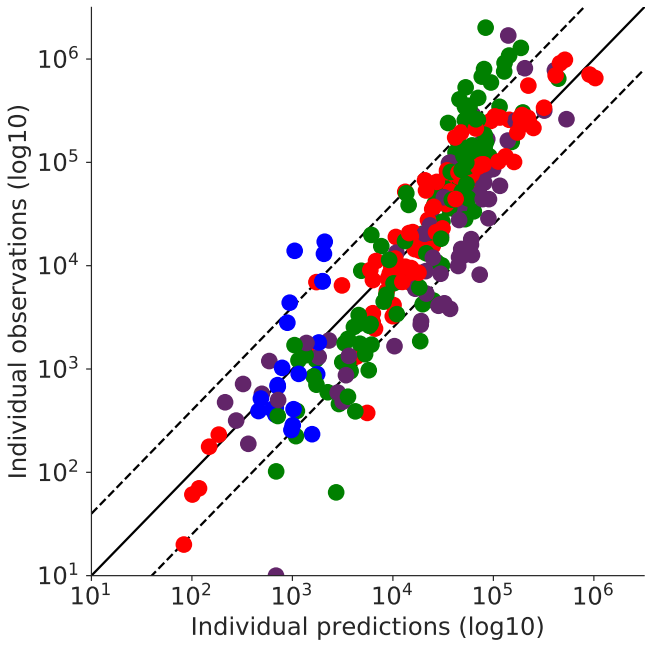



Figure 2: For each CD8 T cell count experimental point, the prediction obtained with System (3) is plotted, for (A) VV data set 1 and (B) Tumor data set 1. Dashed lines represent differences of $\pm$ 0.5 log10 for VV and $\pm$ 0.6 log10 for tumor (maximum estimated error). In both figures, naive (blue), early effector (red), late effector (green), and memory (purple) cell counts are depicted, and the solid black line is the curve $y = x$.

| Parameters | Units | Estimated Values | | Values from |
|---|---|---|---|---|
| | | VV | Tumor | Crauste et al. (2017) |
| | | data set 1 | data set 1 | |
| *Parameters fixed within the population* | | | | |
| $\overline{\delta}_{NE}$ | day$^{-1}$ | 0.0007 | 0.018 | 0.009 |
| $\overline{\delta}_{EL}$ | day$^{-1}$ | 0.10 | 0.11 | 0.59 |
| $\overline{\delta}_{LM}$ | day$^{-1}$ | 0.09 | 0.08 | 0.03 |
| $\overline{\rho}_I$ | day$^{-1}$ | 0.1 | 0.1 | 0.2 |
| *Parameters varying within the population* | | | | |
| $\mu_N$ | day$^{-1}$ | 0.66 | 0.41 | 0.75 |
| $\omega_{\mu_N}$ | day$^{-1}$ | 0.3 | 0.2 | - |
| $\mu_E$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 6.0 | 7.3 | 21.5 |
| $\omega_{\mu_E}$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 0.9 | 0.8 | - |
| $\mu_I$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 2.4 | 2.1 | 1.8 |
| $\omega_{\mu_I}$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 1.0 | 0.6 | - |
| $\rho_E$ | day$^{-1}$ | 1.0 | 0.7 | 0.64 |
| $\omega_{\rho_E}$ | day$^{-1}$ | 0.2 | 0.4 | - |
| *Residual errors* | | | | |
| $a_N$ | cell counts (log10) | 0.2 | 0.5 | - |
| $a_E$ | cell counts (log10) | 0.4 | 0.4 | - |
| $a_L$ | cell counts (log10) | 0.5 | 0.6 | - |
| $a_M$ | cell counts (log10) | 0.3 | 0.6 | - |

Table 3: Estimated parameter values for VV and tumor data sets 1 (median of log-normal distribution for parameters with random effects: $\mu_N$, $\mu_E$, $\mu_I$ and $\rho_E$) and parameter estimation from Crauste et al. (2017) (VV immunization). Estimations have been performed independently.

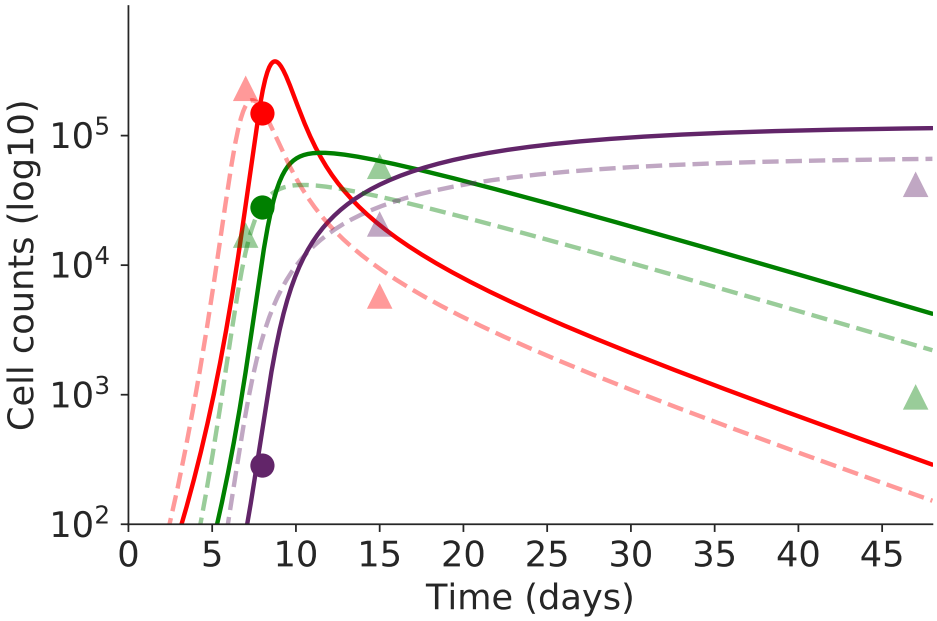



Figure 3: The dynamics of three subpopulations (Early - red, Late - green, Memory - purple) are simulated for two individuals. One individual is displayed as triangles and dashed lines (3 measurements on days 7, 15 and 47pi) and the second as circles and solid lines (only 1 measurement on day 8pi). Although each individual is not characterized by enough experimental measurements to allow parameter estimation on single individuals, nonlinear mixed effects models provide individual fits by considering a population approach.

and Tumor immunizations, and with only few data points per individual.

The estimated parameter values obtained with System (3) for VV or Tumor data sets are in the same range as the estimation previously performed on average cell counts on similar experimental sets (VV immunization, Crauste et al. (2017)), see Table 3. Major differences are observed for estimated values of differentiation rates, yet for the 3 estimations (VV data set 1, Tumor data set 1, Crauste et al. (2017)) parameter values remain in the same order of magnitude, indicationg a good consistency between the two studies. Estimated values of parameter $\overline{\delta}_{NE}$ show the largest relative differences (more than 10-fold from one another). Yet, the largest difference is observed between VV and Tumor data sets obtained with System (3), rather than between these values and the one obtained in Crauste et al. (2017). This may highlight a potential difference in the capacity of the two immunogens (VV and Tumor) to

activate naive cells. This is investigated in the next section.

### 3.3 Immunization-dependent parameters

**Parameter comparison between immunizations.** VV and Tumor induced immunizations differ in many aspects. VV immunizations are virus-mediated, use the respiratory tract to infect cells, and trigger an important innate response. Tumor immunizations involve eukaryotic cells bearing the same antigen, use subcutaneous routes, and induce a reduced innate response. From the independent estimations on VV and Tumor data sets (Table 3), differences between estimated values of fixed effects can be computed. Differences between estimated values are large for parameters – in decreasing order – $\overline{\delta}_{NE}$ (2471%), $\mu_N$ (39%), $\rho_E$ (33%), $\mu_E$ (22%), and $\overline{\delta}_{LM}$ (17%). These large differences may result from biological processes involved in the CD8 T cell response that differ depending on the immunogen.

Consequently, using both data sets (VV and Tumor) as observations may highlight which parameters are required to be significantly different to describe both data sets.

**Parameters depending on immunization.** To perform this analysis, we combine the VV and Tumor data sets 1 and we include a categorical covariate into the model to estimate parameter values (see section 2.3). We are then able to quantify the differences in parameter values obtained when fitting CD8 T cell dynamics using different immunogens (see section 2.3).

The covariate allows to identify parameter values that are significantly different between a response to Tumor and a response to VV. It is a parameter that is added to the fixed effects of the five parameters that showed the larger differences in the initial estimation: $\overline{\delta}_{NE}$, $\mu_N$, $\rho_E$, $\mu_E$ and $\overline{\delta}_{LM}$. This results in the estimation of two different parameter values for parameters $\overline{\delta}_{NE}$ and $\overline{\delta}_{LM}$ (that are fixed within the population) and two probability distributions with different mean values for parameters $\mu_N$, $\rho_E$ and $\mu_E$ (that are allowed to vary within the population).

One may note that adding a covariate increases the number of parameters to estimate. However, the number of parameters is not doubled, since we assumed that parameters without covariates are shared by both immunization groups. In addition, the data set is larger, since it combines VV and Tumor measurements. Hence the number of parameters with respect to

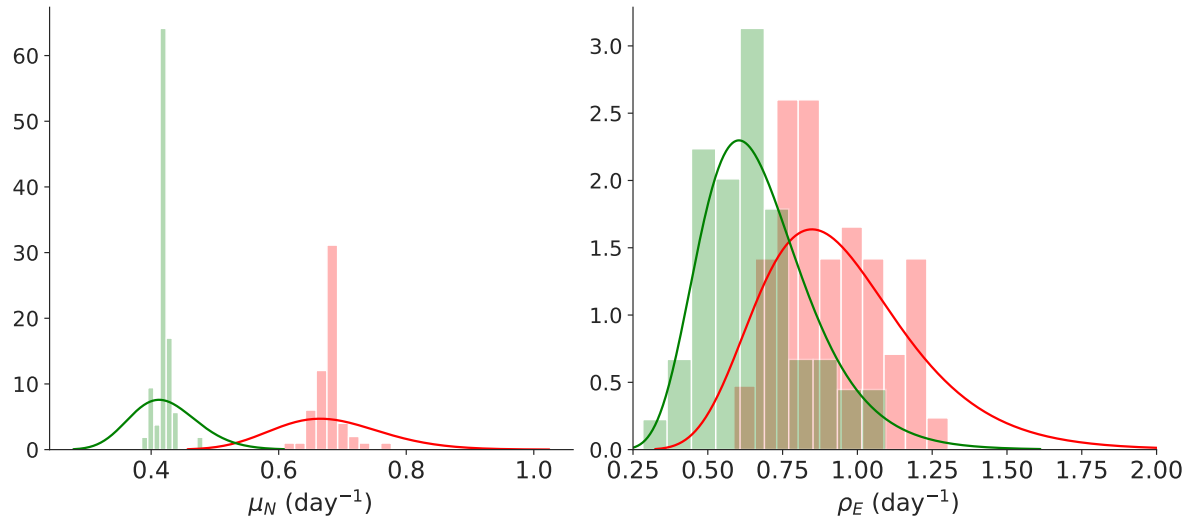



Figure 4: Probability distribution of parameters with a covariate varying within the population, $\mu_N$ (left) and $\rho_E$ (right). Since a covariate parameter has been included, the two estimated distributions are plotted. In addition, histograms of estimated individual parameter values are plotted (green for Tumor, red for VV).

the amount of data remains reasonable.

From this new estimation, we conclude that among the five selected parameters the covariates of only four of them are significantly different from zero: $\overline{\delta}_{NE}$, $\mu_N$, $\rho_E$ and $\overline{\delta}_{LM}$ (Wald test, see Table S3 and section 2.3). The estimation is therefore performed a second time assuming parameter $\mu_E$ distribution is the same in both groups. In these conditions, the estimated covariates for the four parameters remain significantly different from zero (Table 4).

Figure 4 shows the estimated distributions for the two parameters $\mu_N$ and $\rho_E$ that vary within the population and for which we included a covariate. Histograms display the estimated individual parameter values of $\mu_N$ and $\rho_E$. For both parameters, the histograms and the theoretical distributions are in agreement. Moreover, histograms for VV and Tumor appear clearly distinct, which is consistent with the inclusion of a covariate.

It is noticeable that histograms of individual values for $\mu_N$ show shrunk values for both VV and Tumor data sets. Indeed, the empirical standard deviations of individual parameter values are small (VV: $\omega_{emp} = 0.02$ ; Tumor: $\omega_{emp} = 0.01$). Based on this estimation, parameter $\mu_N$ could then be fixed within the population. Yet, when estimating parameter values on VV

| Parameters | Units | VV | Tumor | $p$-value |
|---|---|---|---|---|
| *Parameters fixed within the population* | | | | |
| $\overline{\delta}_{NE}$ | day$^{-1}$ | 0.0010 | 0.0156 | $10^{-15}$ |
| $\overline{\delta}_{EL}$ | day$^{-1}$ | 0.1 | 0.1 | - |
| $\overline{\delta}_{LM}$ | day$^{-1}$ | 0.1 | 0.07 | 0.03 |
| $\overline{\rho}_{I}$ | day$^{-1}$ | 0.1 | 0.1 | - |
| *Parameters varying within the population* | | | | |
| $\mu_N$ | day$^{-1}$ | 0.68 | 0.42 | $10^{-5}$ |
| $\omega_{\mu_N}$ | day$^{-1}$ | 0.13 | 0.13 | - |
| $\mu_E$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 5.9 | 5.9 | - |
| $\omega_{\mu_E}$ | $10^{-6}$ cell$^{-1}$ day$^{-1}$ | 0.9 | 0.9 | - |
| $\mu_I$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 2.6 | 2.6 | - |
| $\omega_{\mu_I}$ | $10^{-5}$ cell$^{-1}$ day$^{-1}$ | 0.8 | 0.8 | - |
| $\rho_E$ | day$^{-1}$ | 0.9 | 0.67 | $10^{-6}$ |
| $\omega_{\rho_E}$ | day$^{-1}$ | 0.3 | 0.3 | - |
| *Residual errors* | | | | |
| $a_N$ | cell counts (log10) | 0.5 | 0.5 | - |
| $a_E$ | cell counts (log10) | 0.4 | 0.4 | - |
| $a_L$ | cell counts (log10) | 0.5 | 0.5 | - |
| $a_M$ | cell counts (log10) | 0.4 | 0.4 | - |

Table 4: Estimated parameter values using combined VV and Tumor data sets 1. Parameters that do not vary within the population are shown in the upper part of the table, whereas individual-dependent parameters are shown in the central part (mean and standard deviation values). Parameters whose values depend on the immunogen (VV, Tumor) are highlighted in grey, and the $p$-value characterizing the covariate non-zero value is shown in the last column.

and Tumor data sets separately, it is not possible to remove the variability of $\mu_N$ which is required to describe the data (see Table 3). We decided to keep $\mu_N$ variable within the population because the model has been validated on synthetic data and on each experimental set separately. Moreover, here we are interested in identifying a potential influence of the immunogen rather than characterizing parameters.

Table 4 gives the estimated values of all parameters in both groups. Regarding parameters that do not vary within the population, values of $\overline{\delta}_{EL}$ and $\overline{\rho}_I$ are the same in both groups, since no covariate is included on these parameters. On the contrary, it is required for parameters $\overline{\delta}_{NE}$ and $\overline{\delta}_{LM}$ to be different to describe each data set, and this difference is accounted for with a covariate parameter.

In summary, we identified parameters whose values are significantly different according to the immunogen used to activate CD8 T cells, these parameters correspond to the dynamics of naive cells ($\overline{\delta}_{NE}$ and $\mu_N$), the proliferation of early effector cells ($\rho_E$), and differentiation to memory cells ($\overline{\delta}_{LM}$). We hence conclude that different immunizations induce variability in the CD8 T cell responses by acting on the first phase of the response (priming, activation of naive cells, expansion of the CD8 T cell population) as well as the development of the memory population.

### 3.4 Predictive immunogen load dynamics

Numerical simulations of System (3) give access to information that may be difficult or impossible to measure, like immunogen load dynamics. For instance, measuring VV load over time in the same mouse is not possible since it requires to kill the animal. Tumor growth, on the contrary, happens locally so tumor volume can be followed for each mouse over time. For individuals with tumor volume measurements (Tumor data set 3, Table 1 and section 2.5), we can qualitatively compare tumor volume evolution with immunogen load dynamics generated by System (3) using the parameter values estimated for each individual.

Figure 5.A shows normalized tumor volumes and simulated immunogen load values for the 20 individuals over 18 days pi, as well as the population behavior (dashed black curve). The population curve peaks at the same time as experimental data, between days 6pi and 8pi. Additionally, the overall dynamics of the population behavior is in agreement with experimental

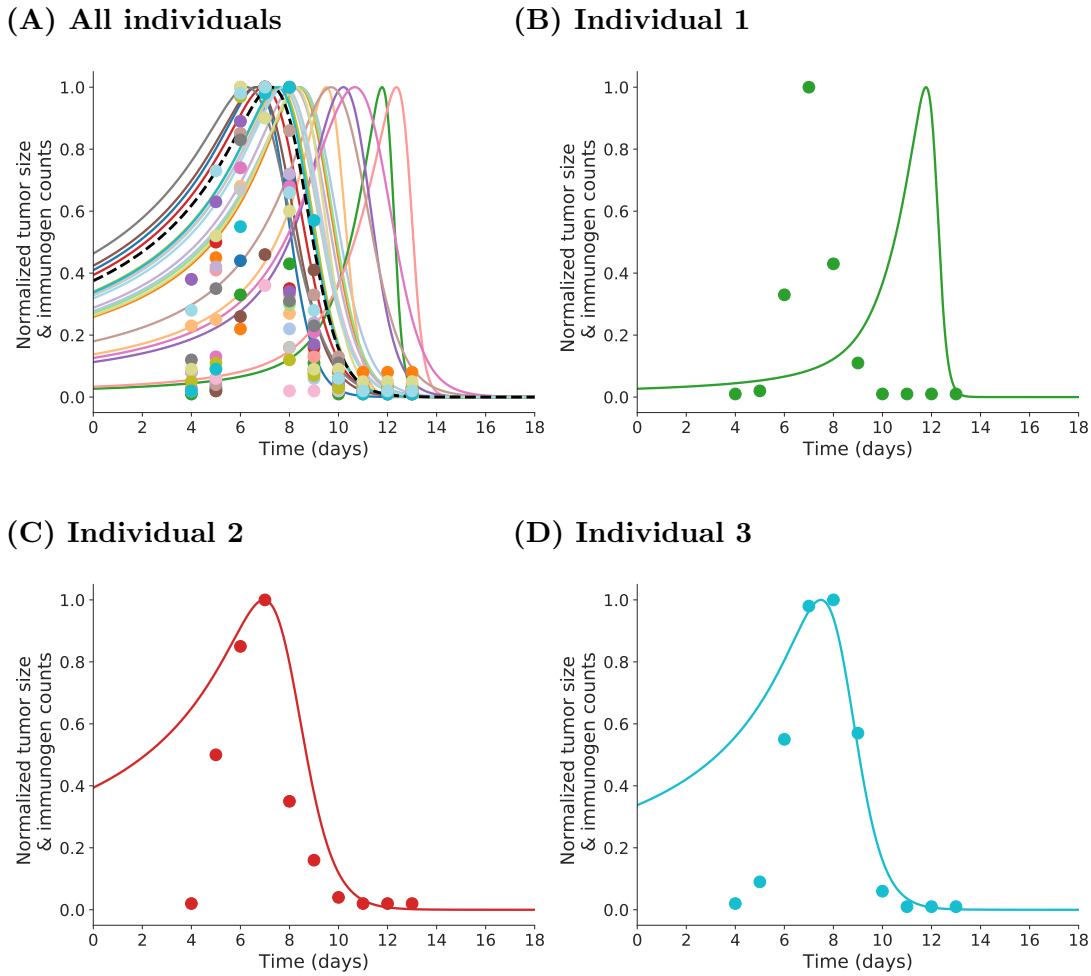



Figure 5: Tumor size volumes and simulated immunogen load dynamics. Both experimental (Tumor data set 3) and simulated values are normalized by their maximal value, in order to compare times at which the maximal value is reached and decay slopes. (A) All individuals are drawn in color (points for experimental data, curves for simulated dynamics) and the population curve in dashed black. (B) to (D) 3 individual dynamics are displayed.

data, except for the slope of the initial increase (but this is due to a difference in values at day 0, see section 2.5). Few simulated curves display dynamics with a peak after day 10pi.

Figures 5.B to 5.D focus on three individuals, with illustrative behaviors. Simulated dynamics in Figure 5.B show a late peak of the immunogen load, after day 10pi, when experimental peak is on day 7pi and the tumor volume has already shrunk by day 9pi. Simulated dynamics are much slower, with a slow increase followed by a rapid disappearance of the immunogen load around day 13pi. On the contrary, Figures 5.C and 5.D show very good agreement between simulated curves and experimental data. Indeed, the time of the peak and the decay rate of the immunogen load for the two individuals are well matched.

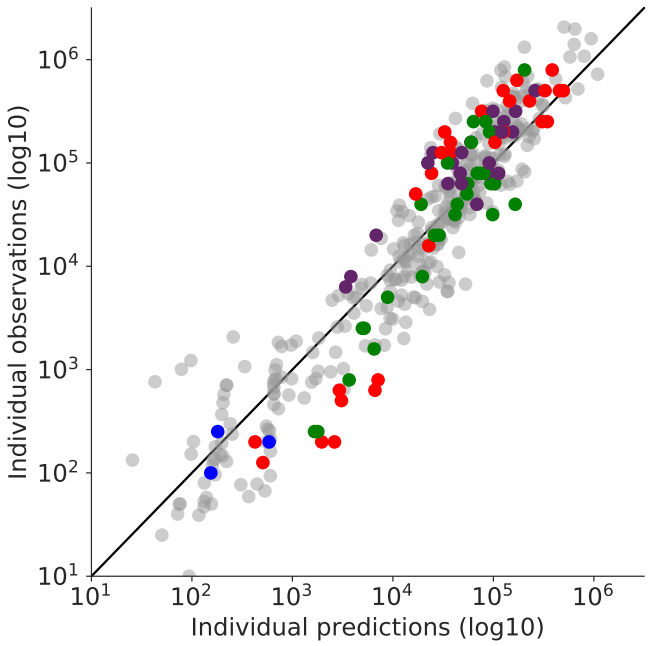
Simulated immunogen load dynamics have been ranked in increasing order according to their ability to reproduce measured tumor volumes (see Supplementary File 2). Half of individual dynamics are well predicted by the model (as illustrated in Figures 5.C and 5.D), whereas one third of individual dynamics are badly predicted (as in Figure 5.B): the peak of immunogen load is strongly delayed in these cases. Good predictions mostly correspond to individuals with CD8 T cell measurements around the time of the immunogen peak, that is between days 6pi and 8pi. This suggests that appropriate measurements of CD8 T cell counts around the time of the peak of the response allow to correctly estimate CD8 T cell dynamics which, in turn, are good predictors of tumor volume dynamics *in vivo*.

### 3.5 Predicting dynamics following VV and Tumor immunizations

To challenge System (3) and the estimated parameters (Table 4), we compare simulated outputs to an additional data set of both VV and Tumor immunizations, VV data set 2 and Tumor data set 2 (Table 1 and section 2.5).

Since we already know the probability distribution of parameters (Table 4), we only estimate individual parameters in order to fit individual dynamics. Results are shown in Figure 6, for both VV data set 2 and Tumor data set 2. It is clear that estimated individual dynamics are consistent with previous individual dynamics estimations. Hence, we validate System (3) and values estimated in both VV and Tumor immunization contexts by showing that estimated parameter values allow to characterize CD8 T cell counts obtained in similar contexts.

**(A) VV**




**(B) Tumor**

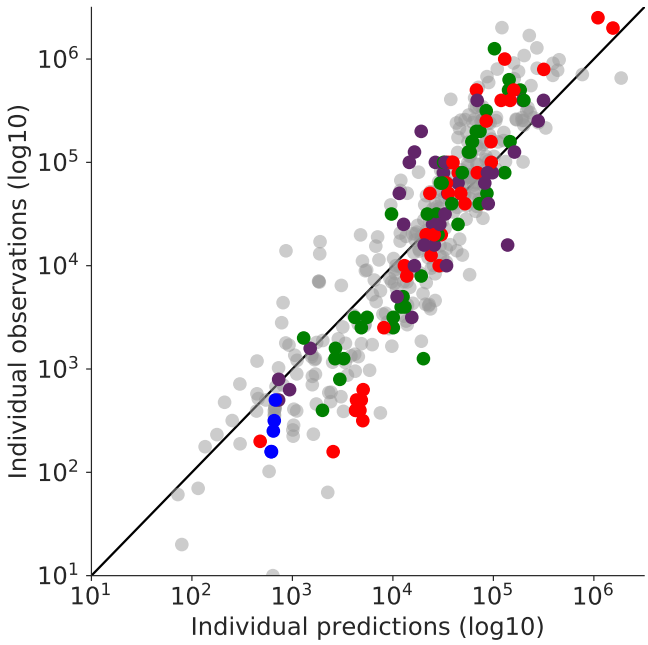



Figure 6: Observations *vs* estimated values of individual CD8 T cell counts, for (A) VV data set 2 and (B) Tumor data set 2. In both figures, naive (blue), early effector (red), late effector (green), and memory (purple) cell counts are depicted, and grey points correspond to individual values from Figure 2. The black straight line is $y = x$.

# 4 Discussion

When following an *in vivo* immune response, experimental measurements are often limited by either ethical issues or tissue accessibility. Consequently, one often ends up measuring cell counts in peripheral blood on a restricted number of time points per individual, over the duration of a response (see Figure 3). With such data, estimation of all model parameters becomes unlikely. Using nonlinear mixed effects models, we propose a dynamical model of CD8 T cell dynamics that circumvents this difficulty by assuming that all individuals within a population share the same main characteristics. The model allows the accurate description of individual dynamics, even though individual measurements are scarce. Indeed, we are able to estimate both good fits and relevant dynamics for individuals with only few cell count measurements, as illustrated in Figure 3. These results indicate that knowledge of population dynamics parameters and numerical simulations complement information given by experimental measurements.

Starting from the model described in Crauste et al. (2017) that could efficiently describe CD8 T cell dynamics, at the level of average population cell-counts present in peripheral blood, we built this nonlinear mixed effects model in a stepwise fashion. The system was first reduced to ensure numerical well-posedness of the parameter estimation. We next identified parameters – hence biological processes – that vary between individuals, and parameters that are fixed within the population. Finally, by adding a covariate to some parameters we identified immunization-dependent parameters.

Noteworthy, from a biological point of view the removal of one parameter (for example, the death rate of late effector cells) does not imply that the corresponding process is not biologically meaningful. However, based on the available data, our methodology found that some processes were negligible in comparison with the ones described by the system's equations. Similarly, we were led to define fixed differentiation rates and immunogen proliferation rate among the population. The constancy of the differentiation rates is in good agreement with the auto-pilot model that shows that once naive CD8 T cells are activated their differentiation in memory cells is a steady process (Kaech and Ahmed, 2001; Stipdonk et al., 2001). Although we cannot exclude that a constant pathogen proliferation rate is due to a lack of data on the pathogen counts, that would have allowed a more refined calibration of immunogen load dynamics,

this result is in good agreement with Crauste et al. (2015) that showed a great robustness of the pathogen proliferation rate. Consequently, inter-individual variability is explained only by variability in mortality rates of all cells (naive, early effector, late effector, and memory) and the immunogen, and proliferation of early effector CD8 T cells. The latter is actually in good agreement with the demonstration that in diverse infection conditions the magnitude of antigen-specific CD8 T cell responses is primarily controlled by clonal expansion (van Heijst et al., 2009). Eventually, using nonlinear mixed effects models we were able to quantitatively reproduce inter-individual variability in two different immunization contexts (VV and Tumor) and provide predictive population dynamics when confronted to another data set (for both immunogens). Therefore, robustness of the model is strong.

The addition of a covariate allowed us to identify parameters that are immunization-dependent. Interestingly they control the activation of the response (priming, differentiation of naive cells, expansion of effector cells) as well as the generation of memory cells. This is again in good agreement with the biological differences that characterize the two immunogens used in this study. Indeed, pathogen-associated molecular patterns (PAMP) associated with vaccinia virus will activate a strong innate immune response that will provide costimulatory signals that in turn will increase the efficiency of naive CD8 T cell activation (Iwasaki and Medzhitov, 2015). In contrast, when primed by tumor cells CD8 T cells will have access to limited amount of costimulation derived from damage associated molecular patterns (Yang et al., 2017). The amount of costimulation will also control the generation of memory cells (Mescher et al., 2006). Hence the addition of covariates to the model parameters has allowed to identify biologically relevant, immunogen-dependent parameters. Nevertheless, using covariates has additional advantages. First, they allow to consider a larger data set (in our case, the combination of two data sets) without adding too many parameters to estimate (4 covariates in our case). This is particularly adapted to situations where only some parameters are expected to differ depending on the data set (here, the immunogen). Second, and as a consequence, data fits may be improved compared to the situation where data sets generated with different immunogens are independently used to estimate parameters. Figure 7 illustrates this aspect: dynamics of two individuals are displayed, with and without covariate. In both cases using the covariate (and thus a larger data set) improved the quality of individual fits,

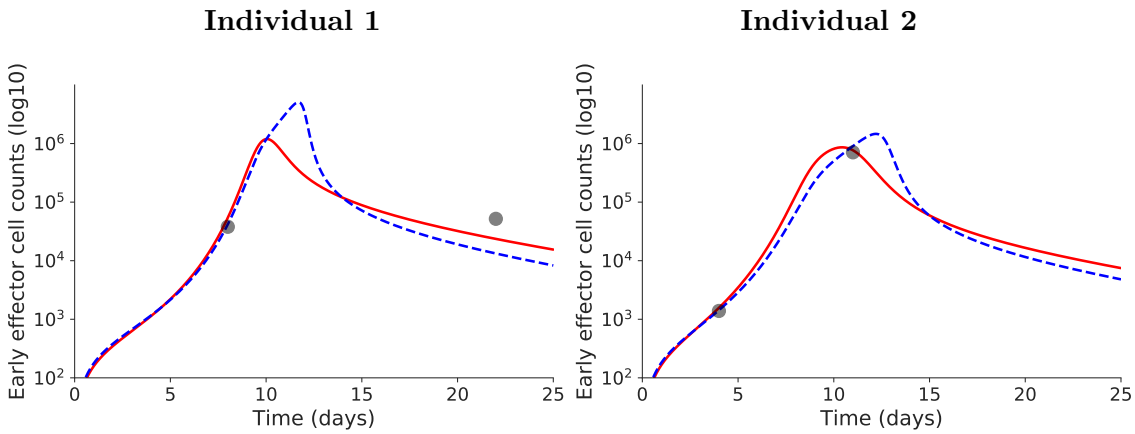



Figure 7: Positive side-effect of using covariates. For two illustrative individuals, accounting for covariates allows to better estimate early effector cell dynamics: red plain curve with covariate, blue dashed curve without covariate.

and generated more relevant dynamics with a peak of the response occurring earlier, before day 10pi. No individual fit has been deteriorated by the use of a covariate (not shown).

The model also proved to be predictive. Indeed, for individuals with tumor volume measurements, the qualitative comparison with immunogen load dynamics was good (section 3.4) even though this immunogen data set has not been considered to estimate parameter values. This reveals that tumor volume evolution may be predicted by CD8 T cell counts in blood samples, provided that CD8 T cell counts are measured around the expected time of the peak of the response (between days 6pi and 9pi). This highlights how measurements of CD8 T cell counts at relevant times (before, around, and after the peak of the response) ensure correct predictions of CD8 T cell dynamics (as already mentioned in Crauste et al. (2017)), and consequently of tumor dynamics. Future works need to focus on this relation, and tumor therapy may benefit from these new findings.

Finally, CD8 T cell response dynamics to both VV and Tumor immunogens were well captured for data sets that had not been used to perform parameter estimation (section 3.5). The behavior of each individual was estimated with the prior knowledge acquired on the population (i.e. fixed parameters values and variable parameter distributions) and proved consistent with previous estimated individual behaviors. The correct prediction of individual behaviors by the model, in a simple mice experiment, paves the way to personalized medicine based on numer-

ical simulations. Indeed, once the population parameters are defined, numerical simulation of individuals can be performed from a few measurements per individual and consequently would allow to adapt personalized therapies.

## Acknowledgements

The authors are grateful to Pr. Adeline Leclercq Samson for sharing her expertise on nonlinear mixed effects models. We thank the BioSyL Federation and the LabEx Ecofect (ANR-11-LABX-0048) of the University of Lyon for inspiring scientific events. This work was supported by Inria PRE MEMOIRE grant and by the ANR predivac grant (ANR-12-RPIB-0011-01). We acknowledge the contributions of SFR Biosciences (UMS3444/CNRS, US8/Inserm, ENS de Lyon, UCBL) and of the CELPHEDIA Infrastructure (http://www.celphedia.eu/), especially the center AniRA in Lyon (AniRA-PBES and AniRA-Cytometrie facilities).

## Competing Interests

The authors have no competing interests.

## References

Antia R, Bergstrom CT, Pilyugin SS, Kaech SM, Ahmed R (2003) Models of CD8 + responses: 1. What is the antigen-independent proliferation program. J Theor Biol 221, 585–598.

Almquist J, Bendrioua L, Adiels CB, Goksör M, Hohmann S, Jirstrand M (2015) A Nonlinear Mixed Effects Approach for Modeling the Cell-To-Cell Variability of Mig1 Dynamics in Yeast. PLoS ONE 10 (4), e0124050.

Althaus CL, Ganusov VV, De Boer RJ (2007) Dynamics of CD8+ T cell responses during acute and chronic lymphocytic choriomeningitis virus infection. J Immunol 179(5), 2944–2951.

Badovinac VP, Haring JS, Harty JT (2007) Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8(+) T cell response to infection. Immunity 26(6), 827–841.

Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JML, Hlatky L, Hahnfeldt P (2014) Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth. PLoS Comput Biol 10(8), e1003800.

Boissonnas A, Combadiere C, Lavergne E, Maho M, Blanc C, Debré P, Combadiere B (2004) Antigen distribution drives programmed antitumor CD8 cell migration and determines its efficiency. J Immunol 173(1), 222–229.

Crauste F, Terry E, Le Mercier I, Mafille J, Djebali S, Andrieu T, Mercier B, Kaneko G, Arpin C, Marvel J, Gandrillon O (2015) Predicting pathogen-specific CD8 T cell immune responses from a modeling approach. J Theoret Biol 374, 66–82

Crauste F, Mafille J, Boucinha L, Djebali S, Gandrillon O, Marvel J, Arpin C (2017) Identification of nascent Memory CD8 T cells and modeling of their ontogeny. Cell Syst 4(3), 306–317.

Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: An overview and update. J Agric Biol Environ Stat 8, 387–419.

De Boer RJ, Oprea M, Antia R, Murali-Krishna K, Ahmed R, Perelson AS (2001) Recruitment times, proliferation, and apoptosis rates during the CD8 + T-cell response to lymphocytic choriomeningitis virus. J Virol 75, 10663–10669.

de Brito C, Tomkowiak M, Ghittoni R, Caux C, Leverrier Y, Marvel J (2011) CpG promotes cross-presentation of dead cell-associated antigens by pre-CD8a+ dendritic cells [corrected]. J Immunol 186, 1503–1511.

Delyon B, Lavielle M, Moulines E (1999) Convergence of a stochastic approximation version of the EM algorithm. The Annals of Stat 27(1), 94–128.

Estcourt MJ, Létourneau S, McMichael AJ, Hanke T (2005) Vaccine route, dose and type of delivery vector determine patterns of primary CD8+ T cell responses. Eur J Immunol 35, 2532–2540.

Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G (2008) Variability and robustness in T cell activation from regulated heterogeneity in protein levels. Science 321, 1081–1084.

Ferenci T, Sápi J, Kovács L (2017) Modelling tumor growth under angiogenesis inhibition with mixed-effects models. Acta Polytechnica Hungarica 14, 221–234.

Ferraro A, D'Alise AM, Raj T, Asinovski N, Phillips R, Ergun A, Replogle JM, Bernier A, Laffel L, Stranger BE, De Jager PL, Mathis D, Benoist C (2014) Interindividual variation in human T regulatory cells. Proc Nat Acad Sci 111(12), 1111–1120.

Fischer A, Rausell A (2016) Primary immunodeficiencies suggest redundancy within the human immune system. Sci Immunol 1, eaah5861.

Grau M, Valsesia S, Mafille J, Djebali S, Tomkowiak M, Mathieu AL, Laubreton D, de Bernard S, Jouve PE, Ventre E, Buffat L, Walzer T, Leverrier Y, Marvel J (2018) Antigen-Induced but Not Innate Memory CD8 T Cells Express NKG2D and Are Recruited to the Lung Parenchyma upon Viral Infection. J Immunol 200(10), 3635–3646.

Iwasaki A, Medzhitov R (2015) Control of adaptive immunity by the innate immune system. Nat Immunol 16, 343–353.

Jarne A, Commenges D, Prague M, Levy Y, Thiébaut R (2017) Modeling CD4+ T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous Interleukin 7. The Annals of Applied Statistics 11(3), 1593–1616.

Jubin V, Ventre E, Leverrier Y, Djebali S, Mayol K, Tomkowiak M, Mafille J, Teixeira M, Teoh DYL, Lina B, et al. (2012) T inflammatory memory CD8 T cells participate to antiviral response and generate secondary memory cells with an advantage in XCL1 production. Immunol Res 52, 284–293.

Kaech S, Ahmed R (2001) Memory CD8+ T cell differentiation : initial antigen encounter triggers a developmental program in nave cells. Nat Immunol 2, 415–422.

Keersmaekers N, Ogunjimi B, Van Damme P, Beutels P, Hens N (2018) An ODE-based mixed modelling approach for B- and T-cell dynamics induced by Varicella-Zoster Virus vaccines in adults shows higher T-cell proliferation with Shingrix compared to Varilrix. bioRxiv, Cold Spring Harbor Laboratory, 348284, doi: https://doi.org/10.1101/348284

Kuhn E, Lavielle M (2005) Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics and Data Analysis 49(4), 1020–1038.

Lavielle M (2014) Mixed effects models for the population approach. Models, Tasks, Methods and Tools. Chapmann and Hall/CRC , 383p.

Li Y, Oosting M, Deelen P, Ricaño-Ponce I, Smeekens S, Jaeger M, Matzaraki V, Swertz MA, Xavier RJ, Franke L, Wijmenga C, Joosten L, Kumar V, Netea MG (2016) Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. Nature medicine 22(8), 952.

Llamosi A, Gonzalez-Vargas A.M, Versari C, Cinquemani E, Ferrari-Trecate G, Hersen P, Batt, G (2016) What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. PLoS Comput Biol 12(2), 1553-7358.

Mescher MF, Curtsinger JM, Agarwal P, Casey KA, Gerner M, Hammerbeck CD, Popescu F, Xiao Z (2006) Signals required for programming effector and memory development by CD8+ T cells. Immunol Rev 211, 81-92.

Miller JD, van der Most RG, Akondy RS, Glidewell JT, Albott S, Masopust D, Murali-Krishna K, Mahar PL, Edupuganti S, Lalor S, Germon S, Del Rio C, Mulligan MJ, Staprans SI, Altman JD, Feinberg MB, Ahmed R (2008) Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines. Immunity 28(5), 710–722.

Monolix version 2018R1. Antony, France: Lixoft SAS, 2018.

Murali-Krishna K, Altman JD, Suresh M, Sourdive DJ, Zajac AJ, Miller JD, Slansky J, Ahmed R (1998) Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. Immunity 8(2), 177-187.

Precopio ML, Betts MR, Parrino J, Price DA, Gostick E, Ambrozak DR, Asher TE, Douek DC, Harari A, Pantaleo G, Bailer R, Graham BS, Roederer M, Koup RA (2007) Immunization with vaccinia virus induces polyfunctional and phenotypically distinctive CD8(+) T cell responses. J Exp Med 204(6), 1405–1416.

Samson A, Donnet S (2007) Estimation of parameters in incomplete data models defined by dynamical systems. Journal of Statistical Planning and Inference 137 (9), 2815–2831.

Stipdonk MV, Lemmens E, Schoenberger S (2001) Naive CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. Nat Immunol 2, 423–429.

Tomayko MM, Reynolds CP (1989) Determination of subcutaneous tumor size in athymic (nude) mice. Cancer Chemother Pharmacol, 24 (3), 148–54.

van Heijst JW, Gerlach C, Swart E, Sie D, Nunes-Alves C, Kerkhoven RM, Arens R, Correia-Neves M, Schepers K, Schumacher TN (2009) Recruitment of antigen-specific CD8+ T cells in response to infection is markedly efficient. Science 325(5945), 1265–1269.

Vetvicka D, Hovorka O, Kovar L, Rihova B (2009) Establishment of imageable model of T-cell lymphoma growing in syngenic mice. Anticancer Res 29(11), 4513-4517.

Villain L, Commenges D, Pasin C, Prague M, Thiébaut R (2018) Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts. Statistics in Medicine 38(2), 221–235.

Wong HS, Germain RN (2018) Robust control of the adaptive immune system. Semin. Immunol. 36, 17–27.

Woodberry T, Gardner J, Elliott SL, Leyrer S, Purdie DM, Chaplin P, Suhrbier A (2003) Prime boost vaccination strategies: CD8 T cell numbers, protection, and Th1 bias. J Immunol 170 (5), 2599–2604.

Xiao Z, Curtsinger JM, Prlic M, Jameson SC, Mescher MF (2007) The CD8 T cell response to vaccinia virus exhibits site-dependent heterogeneity of functional responses. Int Immunol 19, 733–743.

Yang D, Han Z, Oppenheim JJ (2017) Alarmins and immunity. Immunol Rev 280, 41–56.

Youngblood B, Hale JS, Kissick HT, Ahn E, Xu X, Wieland A, Araki K, West EE, Ghoneim HE, Fan Y, Dogra P, Davis CW, Konieczny BT, Antia R, Cheng X, Ahmed R (2017) Effector CD8 T cells dedifferentiate into long-lived memory cells. Nature 552(7685), 404–409.