Analysis of Practical Identifiability of a Viral Infection Model

Van Kinh Nguyen[1,2,3], Frank Klawonn[4,5], Rafael Mikolajczyk[6,7,8,9], Esteban A. Hernandez-Vargas[1]*

1 Systems Medicine of Infectious Diseases, Department of Systems Immunology and Braunschweig Integrated Centre of Systems Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany, 2 Epidemiology Department, Ho Chi Minh University of Medicine and Pharmacy, Ho Chi Minh, Vietnam, 3 PhD Programme "Epidemiology", Braunschweig-Hannover, Germany, 4 Biostatistics, Helmholtz Centre for Infection Research, Braunschweig, Germany, 5 Department of Computer Science, Ostfalia University, Wolfenbüttel, Germany, 6 Epidemiological and Statistical Methods, Helmholtz Centre for Infection Research, Braunschweig, Germany, 7 German Centre for Infection Research, site Hannover-Braunschweig, Germany, 8 Hannover Medical School, Hannover, Germany, 9 [Institute of] Medical Epidemiology, Biometry and Informatics, Martin-Luther University Halle-Wittenberg, Germany

* esteban.vargas@helmholtz-hzi.de

Abstract

Mathematical modelling approaches have granted a significant contribution to life sciences and beyond to understand experimental results. However, incomplete and inadequate assessments in parameter estimation practices hamper the parameter reliability, and consequently the insights that ultimately could arise from a mathematical model. To keep the diligent works in modelling biological systems from being mistrusted, potential sources of error must be acknowledged. Employing a popular mathematical model in viral infection research, existing means and practices in parameter estimation are exemplified. Numerical results show that poor experimental data is a main source that can lead to erroneous parameter estimates despite the use of innovative parameter estimation algorithms. Arbitrary choices of initial conditions as well as data asynchrony distort the parameter estimates but are often overlooked in modelling studies. This work stresses the existence of several sources of error buried in reports of modelling biological systems, voicing the need for assessing the sources of error, consolidating efforts in solving the immediate difficulties, and possibly reconsidering the use of mathematical modelling to quantify experimental data.

Introduction

Mathematical modelling approaches have been playing a central role to describe many different real life applications [1]. In biological and medical research, benefits from mathematical modelling are not only generating and validating hypotheses from experimental data but also simulating these models has led to testable experimental predictions [2]. Presently, quantitative modelling is among the renowned methods used to study several aspects of viral infection diseases from virus-host interactions to complex immune responses systems and therapies [3–6], cancer [7], and gene networks [8].

Mechanistic models aim to mimic biological mechanisms that embody some essential and exciting aspects of a particular disease. These can serve either as pedagogical tools to understand and predict disease progression, or act as the objects of further experiments. However, mathematical models at heart are incomplete, and the same process can be modelled differently, if not competitively. Inherently, a model should not be accepted or proven as a correct one, but should only be seen as the least invalid model among the alternatives. This selection process can be done, for example, by fitting the models to experimental data and comparing the models' goodness-of-fit criterion e.g., [9–11]. Alternatively, modelling can be used as "thought experiments" without considering parameter fitting to experimental data. For instance, on the basis of various simple mathematical models, the hypothesis that the immune activation determines the decline of memory CD4+ T cells in HIV infection can be rejected [12].

Assuming there exists a model that represents properly the problem at hand, model fitting to experimental data is still subject to a number of factors that can distort parameter estimates. For instance, mathematical models are often non-linear in structure and their parameters can be strongly correlated, posing troublesome issues to parameter estimation procedures, e.g., parameter identifiability [13] or parameter sensitivity [14]. These problems aggravate as experimental data are often sparse, scarce, and vary by orders of magnitude [9, 13, 15–19]. Consequently, accurate estimation of biological parameters is inherently problematic, if not impossible [14].

Nevertheless, above issues are not *de facto* worrisome barriers if the primary purpose is to evaluate working hypotheses [14, 20]. In this case, finding a working set or a plausible range of parameter values can be deemed sufficient. However, doubts arise when mathematical models are used explicitly to extract biological meaning of model parameters, which is more evident and likely to happen in those applications where the prediction aspect is inferior. For example, modelling studies on acute diseases with predictable dynamics, such as influenza virus infection [3, 4, 21–24], have been pursuing the quantification of parameters rather than prediction. Additionally, the conventional practice of making use of the formerly estimated parameters in mathematical modelling could propagate the potential inaccurate parameters to the subsequent works, weaken the validity of parameter estimates and consequently reduce the model merit. Several approaches have been offered to provide assessments and understanding on the soundness of the estimated parameters, e.g., [13, 14, 19]. However, in fact, the booming works on mathematical models in biological and medical research over the last years have been accompanied with a disproportionately low amount of assessments on parameter validity (Fig 1). This quantitative recapitulation of biological parameters without a doubt has raised concerns in the scientific community, provoking questions on the appropriateness of the approach.

To keep modelling biological systems from being mistrusted, an important step is to recognize potential sources of error and to assess them in publications. This practice will not only increase the parameter credibility but also save the future works from repeating the same mistakes. In this paper, numerical simulations of a popular mathematical model in viral infection research are used to exemplify various sources of error that exist even when a correct and structurally identifiable model is being used to fit designated data. Thus, this paper provides different angles of the limitations in quantifying biological parameters in mathematical modelling. The results are entrusted to foster rich conversations in the scientific community towards efforts in posing good practice guidelines; and hopefully, to bring changes in the school of thoughts of mathematical modelling.

Materials and Methods

Studying sources of error from real experimental data can be debatable [26] as it limits the abilities to examine a key issue being studied in isolation from confounding factors. Therefore, a
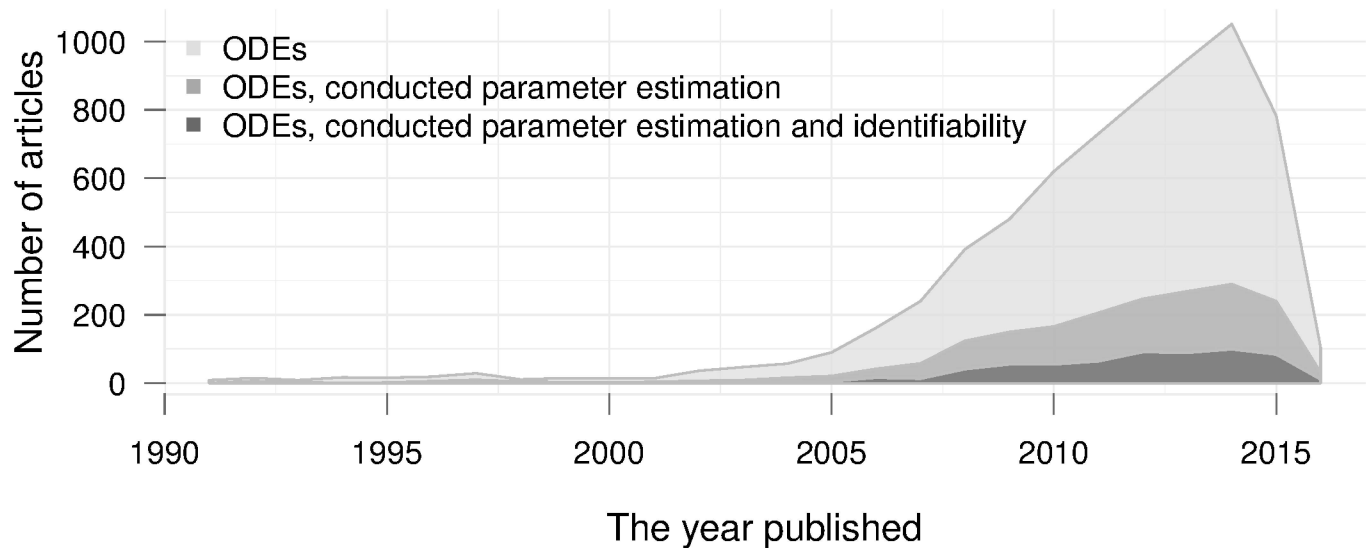
Fig 1. Trends of applied mathematical modelling in biological and medical research over the last decades. ODEs: papers that used ordinary different equations. The data were queried directly from PubMed Central (PMC) database (see S5 Text). The curtailment in the recent years is due to the embargo period of publishing to PMC database [25].

doi:10.1371/journal.pone.0167568.g001

typical mathematical model structure was used to generate synthetic experimental data in various scenarios that resemble data in practice. Each scenario was designed to address a particular source of error while other potentially influential factors were kept under control. Parameter estimation procedures were then applied to recover the model parameters, isolating the impact of each factor on the parameter estimates. The errors of the parameter estimates were evaluated in terms of differences between the estimated values and the original simulation parameters.

To this end, the target-cell limited model presented in [27] was adopted owing to its role as the core component of more than a hundred publications in virus research, e.g., influenza [11, 21, 22, 27–30], HIV [31–34], and Ebola [35] among others [5]. The model includes three compartments: uninfected cells (U), infected cells (I), and viral load (V). It consists of four parameters with partially observable dynamics, i.e., only the viral load dynamics is measured. The model reads as follows

$$dU/dt = -\beta UV, \tag{1}$$

$$dI/dt = \beta UV - \delta I, \tag{2}$$

$$dV/dt = pI - cV, \tag{3}$$

where the parameters to be estimated from experimental data are β, δ, p, and c, which represent the rates of effective infection, infected cell death, viral replication, and viral clearance, respectively. Structural identifiability analyses of the model have shown that the four model parameters can be identified if initial conditions are known [9]. In this model, virus infection is limited only by the availability of the uninfected cells. The uninfected cells are infected by the viruses and become infectious, consequently, these infectious cells are able to release virus particles.

Influenza A virus (IAV) infection dynamics data were considered as reference because of its acute progression nature. It has also been described and modelled extensively [4]. Simulation
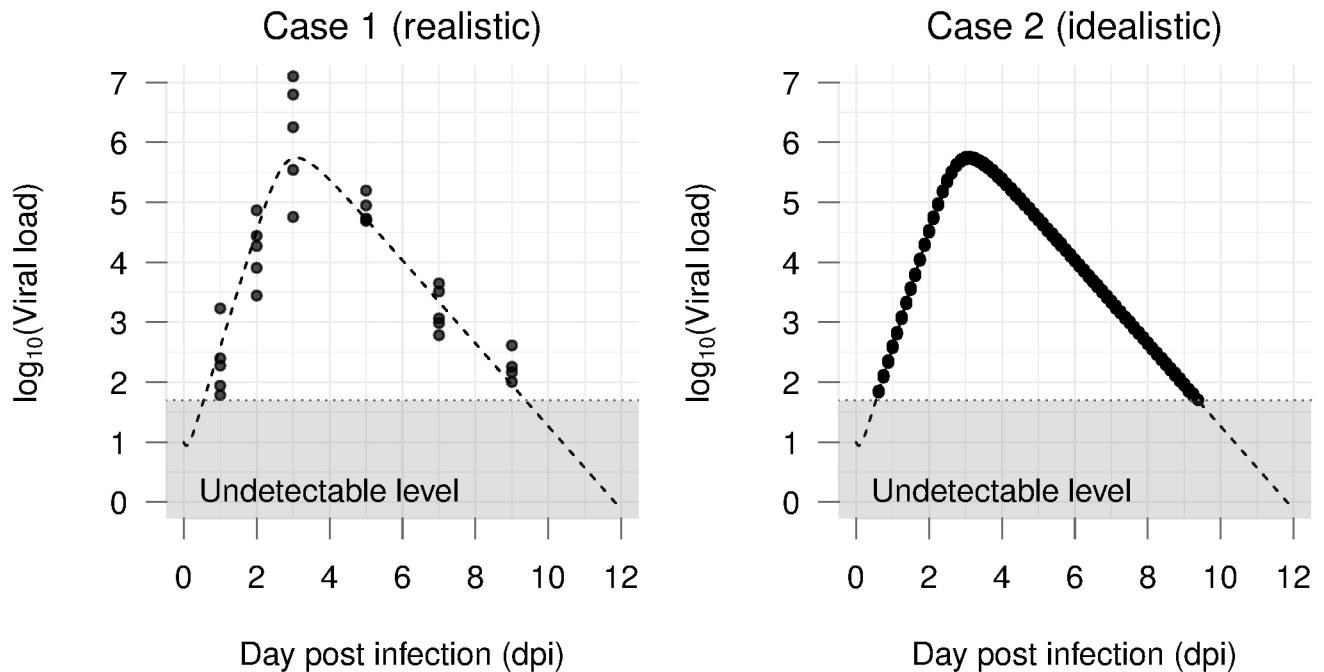
Fig 2. Examples of generated data of iav. Left: A realistic data set with five replicates per time point and measures at irregular time points. **Right**: An idealistic data set with thirty replicates per time point and measures every three hours. The measurement errors are assumed following $N(0, 0.25)$, $N(0, 0.01)$ on log base ten scale for Case 1 and Case 2, respectively.

doi:10.1371/journal.pone.0167568.g002

scenarios were differentiated in number and spacing of sampling time points, number of replicates per time point, and whether a certain experimental setting is known. The reference parameters were chosen to have synthetic data that closely resemble typical experimental observations. In particular, iav infection kinetics comprises a quick viral replication and a peak around the 2nd and 3rd dpi [3, 4, 21]. The viral load stays in a plateau shortly before it declines to an undetectable level below lod at day 7–12 onwards [3, 4, 21]. The half-life of infected cells ($\delta = 1.6$) was assumed shorter than that of healthy epithelial cells [36]. The viral clearance rate ($c = 3.7$) was approximated to that presented in [37]. The remaining two parameters were chosen such that the viral load exponentially increases and peaks at 3rd dpi then declines to the undetectable level at 9th dpi, i.e., $\beta = 10^{-5}$, $p = 5$. Unless stated otherwise, the initial number of uninfected cells, infected cells and virus inoculum are assumed known with $U_0 = 10^6$, $I_0 = 0$, and $V_0 = 10$, respectively. To reflect the lod, generated data points that are smaller than 50 pfu/ml were removed. The measurement error is assumed to be normally distributed on log base ten, i.e., $\sim N(0, 0.25)$ unless otherwise stated. Fig 2 visualizes the two examples of the generated data that mimic a realistic and an idealistic situation. Outlines of the detailed variations of the generated data are shown in Table 1.

To illustrate the error of the recovered parameter estimates, their profile likelihood [13], bias, and variance were computed. In this paper, parameters profiles are used to show if a parameter was identified without error. Each scenario was simulated with a large number of times to compute the bias and the variance [38] of the estimator. Posterior samples [39], weighted bootstrapping samples [40], and the confidence interval based on the approximate observed Fisher information [41] are used to express uncertainty measures where relevant. Simulations were performed using R 3.0.2 [42] and its provided packages when available, otherwise the computations were done with self-written R code.

Results

Choices of initial conditions skew parameter estimates

Generally, the model's initial conditions can be included in parameter estimation procedures as unknown parameters. However, this can be intractable as more nuisance parameters are involved in the estimation routine. An appealing approach is to fix initial conditions and estimate only the model parameters, e.g., [3, 21, 27]. In the target cell model, it is reasonable to assume that the number of infected cells at the beginning of infection is zero. However, the initial number of target cells and inoculum are experiment-dependent and can only be roughly approximated, e.g., initial inoculum had been chosen as half of lod [45].

To isolate the effect of initial conditions on parameter estimates, an idealistic scenario was generated (Fig 2-Case 1, Table 1.a) and the model was fitted with the de algorithm as described in S1 Text. When the correct initial conditions of the model were used, the de algorithm returned the exact values of the parameters (S1 Fig). Afterwards, we varied the initial value of each component in the model and re-estimated the parameters and the corresponding profile likelihood. Because the data and the estimation algorithm did not changed, any error emerging in the parameter estimates can be attributed solely to the changes in the initial conditions. The tested values of the initial conditions were chosen to reflect biological range and its uncertainty in practice, i.e., the initial viral load were under the lod while the number of the initial target cells were about 10^7 (Fig 2). Simulation results show that any deviations, even those that seem negligible, result in erroneous parameter estimates (Fig 3). Arbitrary choices of the initial viral load lead to very different estimates of essential viral infection kinetics parameters, e.g., both the replication rate p and the viral clearance c were estimated six times higher than the true value when fixing the initial viral load at half lod. Furthermore, subjective choices of model initial conditions result in flatness in the profile likelihood and shift the minimum in the parameters space. Note that the cost function values, i.e., rmse in this case, of different parameter sets are still almost identical.

Table 1. Summary of different simulation settings.

Illustrating error in	Time scheme*	Replicates**	Noise***	Initial condition	Algorihtms
a. Initial conditions	t3	30	N(0,0.01)	Vary	DE [43]
b. Measurement error	t3	30	Vary	Fixed	DE [43]
c. Effect of time scheme	Vary	10	N(0,0.25)	Fixed	L-BFGS-B [44]
d. Effect of number of replicates	t24	Vary	N(0,0.25)	Fixed	L-BFGS-B [44]
e. Algorithms	tn1	5	N(0,0.25)	Fixed	Vary
f. Robust estimation	t24	10	N(0,0.25)	Fixed	Vary
g. Polynomial estimate of V_0	Vary	Vary	N(0,0.25)	Not used	GAM [38]
Parameters	Reference	Lower bound+	Upper bound+		
β	10^{-5}	10^{-7}	10^{-3}		
δ	1.6	10^{-2}	10^2		
p	5	10^0	10^2		
c	3.7	10^{-1}	10^2		
Initial condition	$U_0 = 10^6$	$V_0 = 10$	$I_0 = 0$		

*t3 to t24 indicate regular sampling time in every 3 to 24 hours; tn1 indicates sampling time similar to that of [11], i.e., at day 1, 2, 3, 5, 7, 9, tn2 is t24 without measurement at the peak day (3rd) and the endpoint (8th);

**per time point;

***on log 10 scale;

+for Differential Evolution (DE) algorithm and L-BFGS-B [44].

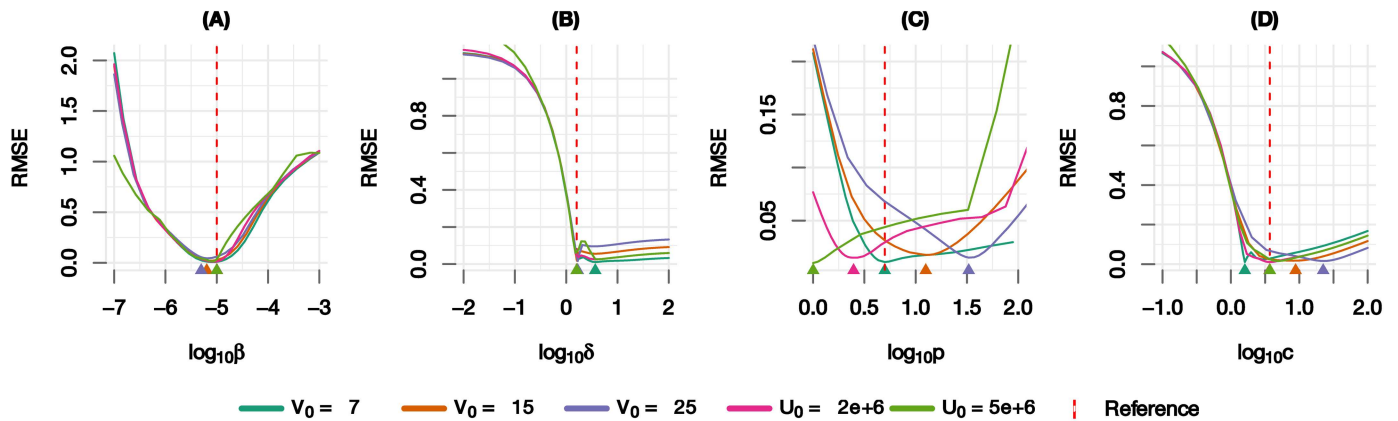doi:10.1371/journal.pone.0167568.t001

Fig 3. Parameters profiles in different initial conditions. The vertical dashed lines indicate the reference values, the triangles are the point estimated for the corresponding initial condition. The data include thirty replicates per time point with minimal measurement error, samples are collected in every three hours during 12 days of the experiment.

Smoothing methods provide efficient ways to illustrate data dynamics. More importantly, they are free from the straitjacket of distributional assumptions and model structure [46]. As such, smoothing techniques could be potentially useful in providing estimates for missing data points and assisting the model parameter estimation. In this context, we tested the General Additive Model (GAM) [38, 47] to extrapolate the initial viral load value in numerous scenarios with a thousand simulations each. Each scenario was a combination of a sampling time scheme and a number of replicates per time point (Table 1.f). Simulation results show that GAM returns estimates for V_0 ranging from 4 to 25 (S3 Fig). Although the range of estimated values is close to the reference value ($V_0 = 10$), this approach may not sufficiently contribute to solving the problem of initial conditions and the subsequent error in parameter procedures.

Measurement error impairs the parameter estimates

The previous scenario assumed virtually no measurement error whilst experimental data in practice can often be seen to vary in one to three orders of magnitude, e.g., see [11, 21, 48]. This is a fundamental barrier in learning from experimental data which will likely not be overcome in the near future. In this context, varying the imposed measurement error of synthetic data and profiling the parameters can provide useful hints on the possible influence of measurement error on parameter estimates. Using the correct initial conditions, an idealistic data set with a large number of replicates and dense sampling time was developed to isolate the effect of measurement error on the accuracy of parameter estimates (Table 1.b). Parameter estimation was done with the de algorithm as it had recovered the exact parameters when there were minimal measurement errors (S1 Fig). As expected, the results show that the estimates drift away from the reference value as the measurement errors increase (S2 Fig). More importantly, increasing the measurement error brings in more flatness to parameter profiles, implying difficulties in estimating parameter values with accuracy.

Data asynchrony generates overlooked measurement error

Experimental data are usually taken for granted as it is observed timely. However, due to different reasons, e.g., host factors, experimental settings, measurement techniques, the observed data are likely to become *asynchronous*. In such situations, measurements recorded at one time point are at best an aggregate measure of the neighbour period.
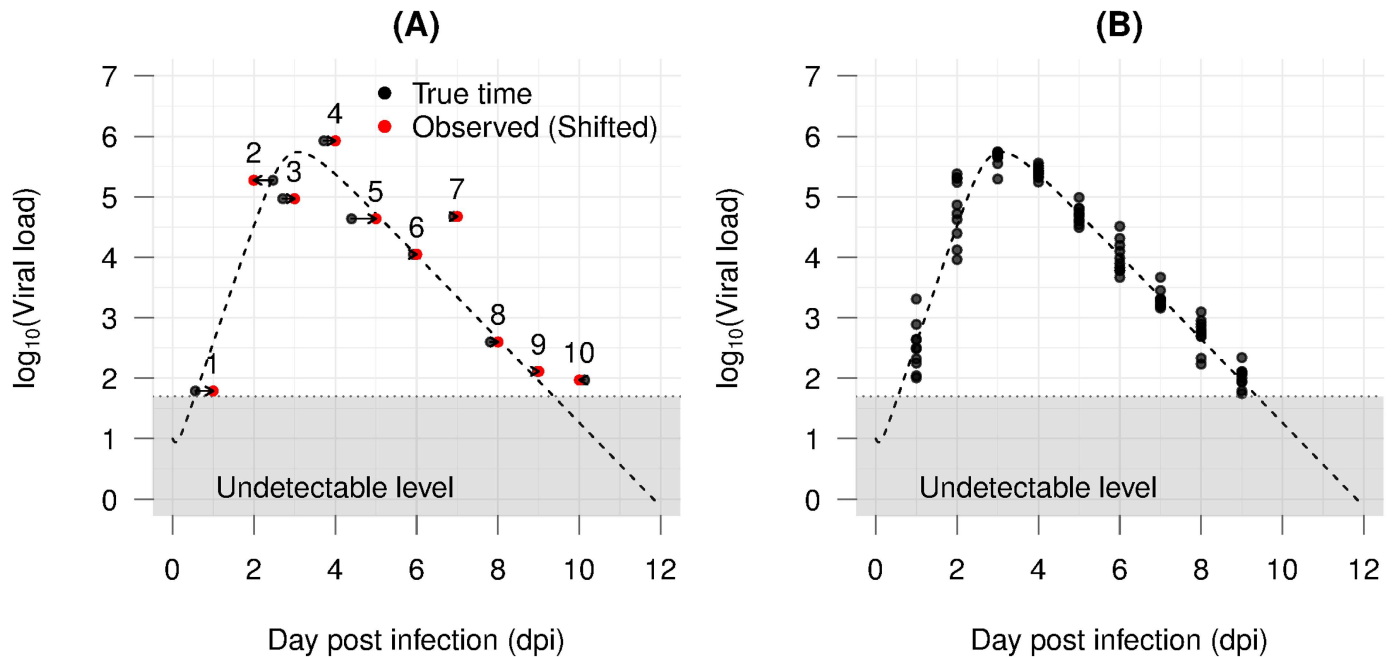
Fig 4. Examples of asynchrony problem. The dotted curve is the true kinetics, dots are data points. (A) Ten subjects with measurement error and asynchrony; red points are the observed data; black points are the actual time the measurement should have represented; arrows indicate directions of time shift; (B) Generated asynchrony data in the absence of measurement error.

The problem stated above can be coined the same as Berkson error type in statistics, where reported measurements are deviations of unknown true values [49]. Let t denote the real time point when an observed value actually happens; the time point will be instead recorded as $t^* = t + v, v \sim N(0, \sigma_t^2)$, where v is an unknown time shift. This shifting could be attributed to the differences in subject-specific responding time to a stimulus and its strength, e.g., infection and the virulence of a virus strain. Furthermore, it is practically not feasible to infect all experimental subjects at the same time or to harvest all required replicates at once. For instance, subject one (Fig 4-A) for unknown reasons has ten hours delay in response to infection, but it is harvested at day first in the *experimental time scale*. This shifting underestimates the viral load on the first day by one order of magnitude.

By and large, the asynchrony leads to data with significant variations even in the absence of measurement error (Fig 4-B). To our knowledge, the scope of this problem has not been addressed or discussed in mathematical modelling works.

Simulations and extrapolation approaches had been shown to be an intuitive way to cope with Berkson error type in bioassay non-linear models [50], and seems to be a promising approach in dealing with data asynchrony in this context. We adapted the approach in an attempt to account for the data asynchrony problem as presented in S4 Text. However, applying simulation steps resulted in a non-linear pattern of parameter estimates (S4 Text). Consequently, it is not possible to make a proper extrapolation to estimates where the parameters are theoretically free from data asynchrony. This result opens a new challenge in mathematical modelling especially for biological processes with fast dynamics.

Sparse and scarce data aggravate the parameter estimates

Data sampling in experimental settings is laborious and costly, consequently, experimental data are usually sparse and scarce. In iav, for example, the viral load had been measured in

time intervals as short as eight hours [48] up to four days apart [51]. In each time point, there could be only a single replicate [52, 53] or up to sixteen replicates in rare cases [51]. In this context, investigating the impact of different data sampling schemes on parameter estimates provides further angles on how prone to error the parameter values can be in practice.

To explore the impact of the sampling time scheme in more practical settings, scenarios in which the data are collected at various time intervals were generated, assuming ten replicates per time point with realistic measurement errors (Table 1.c). For each scenario, the estimation using the de algorithm was repeated on a thousand generated datasets to calculate bias and variance. Unsurprisingly, the parameter estimates variance increases as the sampling time gets sparser (Fig 5A–5D). Bias-variance of both parameters p and c escalates from 16h time scheme to 20h time scheme. It can be observed in Fig 5 that more precision is obtained by increasing the sampling time frequency, but little is gained in parameter accuracy. The practical experiment sampling time tn1 (6 time points and 5 replications) and tn2 (sampling every day but missing the peak day and last time point) manifest the highest variance (Fig 5).

Analysis of the number of replicates per time point was also examined in a similar fashion. The every day sampling scheme (t24) was chosen to see if the estimates could be improved by changing the number of replicates per time point (Table 1.d). As expected, the variance of most parameters increases when the number of measurements per time point is reduced, as well as the bias (Fig 5E–5H). There is not much gain in parameter accuracy as the number of replicates per time point roughly doubles from 13 to 30, and even less so the parameters β and δ.

Experimental design exhibits a strong influence on the magnitude of the bias, i.e., varying either the sampling time or the number of replicates per time point yield steep changes in bias magnitude. It seems that experimental design rewards more accuracy in parameter estimates by sampling more frequently rather than sampling more extensively (Fig 5).
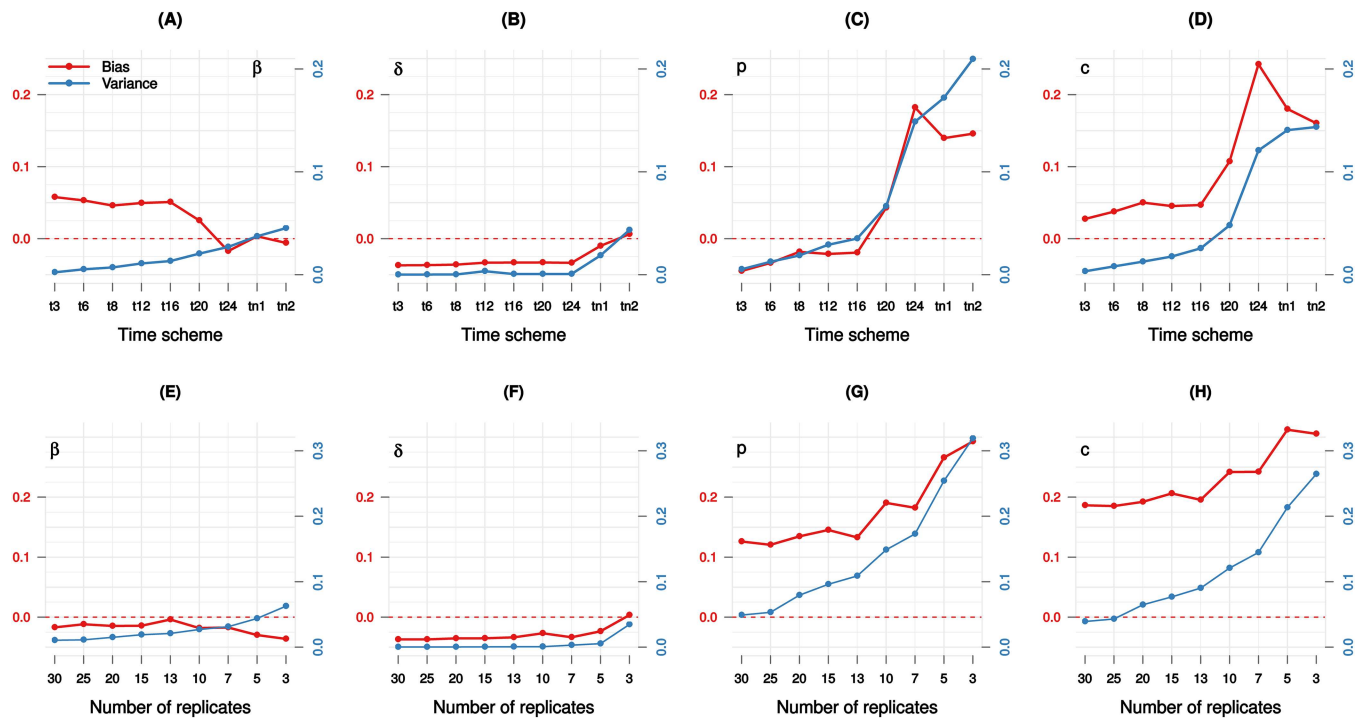

Fig 5. Sampling time scheme and parameter bias-variance. Top: t3 to t24 indicate regular sampling time in every 3 to 24 hours respectively; tn1 indicates sampling time similar to that of [11]; tn2 is t24 without measurement at the peak day (3rd) and the endpoint (8th). Ten replicates were generated per time point. Each scenario (point) is the result of 1000 simulations. **Bottom**: The every day sampling scheme (t24) is used.

doi:10.1371/journal.pone.0167568.g005

It is worth noting that the examples thus far have shown that some model parameters were always estimated with larger error than the others given the same context (Figs 3, 5 and S2 Fig). This underlines the discrepancy of parameter accuracy level in a model, i.e., differences of parameter roles in a model lead to different levels of parameter accuracy.

Parameter estimation algorithms agree on erroneous estimates

Thus far the de algorithm has shown that it can recover the exact parameters in an idealistic case. However, using a wrong assumption for the model or poor experimental data can both drive the algorithm to the wrong estimates. With the progress in parameter estimation techniques these days, it would be intriguing to know whether there is an algorithm that can recover more accurate estimates given a realistic experimental data set. It has been shown that there is not a definitive algorithm that is recommended for all situations [54]. To this end, we challenged the state-of-the-art algorithms to estimate the four model parameters in a practical context (Fig 2-Case 1). We assume that the initial conditions are known to make the parameters structurally identifiable [9] (Table 1.e). The tested algorithms include a box-constraint local estimation algorithm, i.e., L-BFGS-B) [44], a global stochastic algorithm i.e., Differential Evolution (DE) [43], and a Bayesian sampling approach, i.e., Adaptive Metropolis-Hasting (MH) [55]. Maximum likelihood estimation (MLE) and least square (LS) estimation were considered where relevant to reflect their usage in practices. To avoid imposing subjective bias in favour of one algorithm, the algorithms were supplied with their recommended settings (S1 and S2 Texts).

Simulation results show that the algorithms return relatively similar, whilst inaccurate point estimates of the parameters, e.g., the two parameters p and c are distanced by one order of magnitude from the correct ones (Fig 6). Interval estimates of the three algorithms cover similar ranges that spread up to two orders of magnitude. Good interval estimates can be obtained only for the parameter δ. Note that the bootstrapping samples exhibit multi-modal curves, implying there there is at least one outlier in the used data [56]. Given the same advantages, the intensive estimation processes do not reward better estimates, i.e., the L-BFGS-B converged in a few minutes, the mh took twenty minutes to converge to a stable distribution (S2 Text) and the bootstrapping with the de algorithm took a day on a high-performance computer with parallel mode enabled (S1 Text).
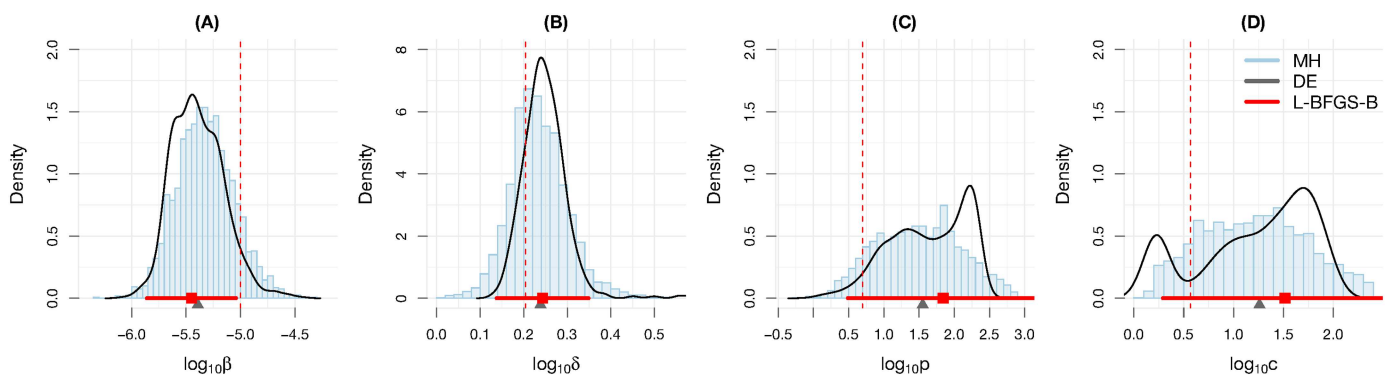

Fig 6. Parameter estimates using different algorithms. The data are collected similar to [11] at day 1, 2, 3, 5, 7, 9, five replicates per time point. **MH**: Metropolis–Hasting estimates are presented as blue histograms of posterior samples; **L-BFGS-B**: local optimization and confidence intervals estimated with approximate Fisher information are presented with red square points and bars; **DE**: point estimates of the global optimization algorithm Differential Evolution and weighted bootstrapping samples are presented with black triangles and curves. The vertical red dashed line is the reference value.

doi:10.1371/journal.pone.0167568.g006

The estimations above were conducted in very favourable conditions that are unlikely to occur in practice, i.e., correct model of the observed data, known initial condition values, real parameter values are in the ranges of searching parameter spaces. Even with those advantages, the parameter values are risky to be interpreted, e.g., the viral replication rate (p) can be recognized as much as one hundred copies per day with strong supports from the interval estimates of different algorithms. Certainly, expecting more accurate estimates in practice are rather overoptimistic, especially when most often none of the above advantages exist in practice. Noting that the same settings for the de algorithm has been shown to be able to recover the exact parameters in an idealistic data set (S1 Fig) and the mh algorithm had converged to a stable distribution (S2 Text). In addition, a test for different combinations of de's algorithm parameters was conducted showing that tuning these DE algorithm's parameters did not improve the rmse (S4 Fig). Thus, it can be inferred that the main force that has driven algorithms to the wrong estimates was the poor experimental setting.

This example also suggests an influential role of outliers in estimating mathematical model parameters. Sources of outliers data could be merely originated from measurement errors or, in fact, come from a different generating mechanism. Robust correction is a popular approach to scale down the influences of those data points by assigning them lesser weights in the optimization process and is believed to reduce their impact on parameter estimates [57]. To test this approach in a model based on ode, thousands of synthetic data sets with outliers were generated. Then the estimation procedures with and without correction for outliers were applied to each of the data sets (S3 Text, Table 1.g). The results show that correction for outliers in this model did not improve the accuracy of the parameter estimates (S3 Text) while the computational cost increased as several rounds of optimization may be needed to achieve stable weights in robust estimation.

Discussion

Mathematical modelling research comprises several disciplines with mutual concepts, approaches, and techniques. Significant concerns about the validity of parameter values had emerged when a majority of modelling works failed to mention the risks of errors in the parameter values while trying to extract biological meaning from it. Throughout this paper, by presenting numerical examples and addressing their implications, we explored how error-prone the parameter estimates are in everyday practices.

It is a catch-22 situation when the way we express a parameter in the model governs its own accuracy. The results show that given the same context, there are model parameters which can be estimated with less error while the others cannot. This reflects the reality that the corresponding component of a parameter in the model is less important in the dynamics. To provide assessments in this aspect, one can head for a global sensitivity analysis approach [58, 59]. However, in biological systems, one can expect joint effects of parameters on the system's output [14]. This leads to situations in which an inaccurate set of parameters generates the same dynamics as the correct one. In this case, the estimated parameter values need to be taken with caution as they are a possible solution in a pool of model solutions. Altogether, in the worst case, the model structure alone already hinders the parameter accuracy, making the interpretation of the parameter values questionable. Reporting assessments of model parameters sensitivity, parameters correlation and identifiability, e.g., [14, 35], can avoid the model results to be over-interpreted.

Leaving out the validity of a model to represent a certain phenomenon, parameter estimation of a correct model with known parameters and fabricated data exhibited several shortcomings. Each algorithm has *per se* a number of parameters as well as configurations that are

required to be specified by the user. The way we choose these settings may directly affect the estimation results. For instance, algorithms might be trapped in a local optimum and this can be tested by tuning the algorithm. Although assessing this aspect is beyond the scope of the work proposed here, simulation results support that the recommended setting is among the best ones for the DE algorithm (S4 Fig). That is why all the algorithms configurations were set based on the literature recommendations as described in S1, S2 and S3 Texts. In addition, testing different algorithms for a study case can help to find out if one algorithm was trapped in a local minimum. However, even in ideal conditions, our results show that different algorithms returned erroneous estimates. Note that the model parameters in the tested situation were set to be theoretically identifiable. This result poses a big question mark on the validity of mathematical model parameters in practice, where the models are more complex but the known model conditions and experimental data are limited.

In accordance with [54], we observed that there is not a conclusive method for parameter estimation. Advanced Bayesian sampling techniques have been developed recently [60], however, these are neither practical to implement in complex mathematical models nor beneficial for parameter accuracy but can serve to explore the posterior distributions. In this family, the Adaptive mh algorithm [55] seems more practical and less dependent on model complexity but it lacks dedicated resources for mathematical modelling (S2 Text). The bootstrap estimates exhibited its shortcoming when the data contained outliers and required more work in modelling context. An example of tackling the outliers based on robust correction approaches also did not show any improvement in parameter accuracy compared to the least square estimation (S3 Text). By and large, one should not expect the power of innovative fitting algorithms to handle sparse and scarce data and to return trustworthy estimates.

Parameters have been deemed reliable when parameter profiles exhibit convex shape [13, 35]. However, our results show that reaching the minimum of the parameter profile curve does not imply that the estimate is accurate (Fig 3). The results show that the revealed parameters are noticeably different from the simulated values regardless that the optimization algorithm has successfully found the minima on the parameter profile curves. Thus, one should take this convergence criterion with caution when attempting to extract biological information, even if one's favourite algorithm works.

In experimental settings, several sources of errors exist that are difficult to overcome. In the target cell model, for example, it is not an easy task for experimentalists to define the exact initial conditions even with the advance of laboratory devices. In contrast, mechanistic models work with a relatively high precision level. This is a straitjacket for all modellers. Working with assumed ode model initial conditions might be the first choice and could be a valid one if the aim is not biological parameter quantification. This approach reduces the number of parameters to estimate in the shortage of data and may hold up in model prediction contexts. However, our results show that the estimates went wrong with seemingly reasonable choices of initial conditions in an idealistic experimental data set (Fig 3). Reporting biological parameters from a mathematical model should incorporate the uncertainty assessment regarding the initial conditions used, for example, a sensitivity analysis with respect to initial conditions.

Smoothing techniques are attractive in mathematical modelling as they can provide a generic portrayal of the dynamics without knowing the underlying mechanism. For instance, the fractional polynomial model was used to estimate the initial viral load [35]. The generalized smoothing approach and local polynomial regression have also been proposed for estimating parameters [61, 62]. However, there is a caveat observed in this paper when modelling with extrapolated data that is most likely to exist in others. Our example of using smoothing techniques to estimate the initial viral load showed that the sparser data tend to have the lesser bias in extrapolating the initial viral load (S3 Fig). All sampling time schemes that collect data

regularly in less than 24 hours exhibit a consistent positive error and *vice versa*. This can be attributed to the fact that all the measurements below lod are removed when generating the data to reflect the undetectable level in practice. The loss of information from those data points under the lod in combination with measurement errors makes it more likely that the observed data at the early, as well as the later time points, of the dynamics could be observed higher than it is. Accordingly, making extrapolation to the time point zero, or to the latest time points can result in an overestimation. This result suggests that fitting the model to a smoothed line could potentially skew the parameter estimates.

We introduced an abstraction of the *data asynchrony* and studied its impact on experimental data which, ultimately, affects the parameter estimates. To the best of our knowledge, this issue has been widely ignored in mathematical modelling. Segregating the measurement error and the errors induced by asynchrony seems impossible. A logical approach is to simulate the amount of time shift and examine its effect on model parameters. In S4 Text we presented an attempt to deal with data asynchrony by simulation and extrapolation to facilitate the conversations in the field. However, further efforts are needed to resolve or, at least, to account for the impact of the asynchrony in mathematical modelling results.

Simulation results show that insufficient data lower the quality of the parameter estimates. In contrast, it is not necessary to collect data extensively, but rather rationally to reduce the error. An experimental design approach suggesting that accurate parameter estimation is achievable with complementary experiments specifically designed for the task [63, 64]. A somewhat simpler version of this idea is that reducing error in parameter estimates can be achieved by fully measuring the model components [23]. Altogether, this supports the fundamental idea as shown in our examples: the primary source to improve the quality of parameter estimates is additional data, in particular, collected in a way designated for a modelling purpose.

The scientific community can actively benefit from an advanced understanding of the sources of error. The purpose of spotting these sources is to avoid or at least to recognize them. Efforts in dealing with errors in parameter estimation shall be documented in publications with mathematical models to strengthen and support further development in the field. There are rooms for further refinement of estimation methodologies to minimize the risk of reporting erroneous estimates and different types of measurement errors in experimental studies. Above all, however, there is a need to unify the efforts in modelling practices, such as develop good practice guidelines for reporting mathematical model assessment results, communicate the difficulties, and construct a common software library. Ultimately, a reassessment of the current approach in modelling experimental data might be needed to shape the future research on a more solid path, maintaining trust in the scientific community of the diligent works in modelling biological systems.

Supporting Information

S1 Text. Limited-memory BFGS (L-BFGS-B) and Differential Evolution (DE) algorithm settings.
(PDF)

S2 Text. Implementation of Adaptive Metropolis-Hasting in ODE models.
(PDF)

S3 Text. Implementation of robust estimation in ODE models.
(PDF)

S4 Text. Implementation of simulation and extrapolation approach.
(PDF)

S5 Text. Description of search results from the PubMed Central database.
(PDF)

S1 Fig. Parameters profiles in an idealistic scenario. The vertical dashed lines indicate the reference values, the solid circles are the estimated values from Differential Evolution algorithm. The data include thirty replicates per time point with minimal measurement error, i.e., $\sim N(0, 0.01)$ on log base ten, samples are collected in every three hours during 12 days of the experiment (Fig 2-Case 2).
(EPS)

S2 Fig. Parameter likelihood profiles in different imposed measurement error. The solid triangles indicate the reference values, the circles are the estimated values corresponding to each condition. Parameters are presented in log base 10 scale. The imposed errors are approximately one to two orders of magnitude of viral load measurements. The profiles lines are standardized by their imposed measurement error for visual purposes.
(EPS)

S3 Fig. Bias-variance of V_0 estimates in different scenarios. Each scenario (point) is the result of 1000 simulations. The legends t3 to t24 indicate a regular sampling time every 3 to 24 hours; tn1 indicates sampling time similar to that of [11]; tn2 is t24 without measurement at the peak day (3rd) and the endpoint (8th).
(EPS)

S4 Fig. RMS values versus F and CR parameters combination of de algorithm. Colours indicate the corresponding values of CR. The plus sign indicates the value that is used in this paper.
(EPS)

Author Contributions

Conceptualization: EHV.

Formal analysis: VKN EHV.

Investigation: VKN.

Methodology: VKN FK EHV.

Project administration: EHV.

Supervision: EHV.

Validation: VKN FK RM EHV.

Visualization: VKN FK RM EHV.

Writing – original draft: VKN FK RM EHV.

Writing – review & editing: VKN FK RM EHV.

References

1. Gershenfeld NA. The Nature of Mathematical Modeling. Cambridge University Press; 1999.
2. Tomlin CJ, Axelrod JD. Biology by Numbers: Mathematical Modelling in Developmental Biology. Nature Reviews Genetics. 2007; 8(5):331–340. doi: 10.1038/nrg2098 PMID: 17440530

3.   Beauchemin C, Handel A. A Review of Mathematical Models of Influenza A Infections Within a Host or Cell Culture: Lessons Learned and Challenges Ahead. BMC Public Health. 2011; 11(Suppl 1):S7. doi: 10.1186/1471-2458-11-S1-S7 PMID: 21356136

4.   Boianelli A, Nguyen VK, Ebensen T, Schulze K, Wilk E, Sharma N, et al. Modeling Influenza Virus Infection: A Roadmap for Influenza Research. Viruses. 2015; 7(10):2875. doi: 10.3390/v7102875

5.   Perelson AS, Guedj J. Modelling Hepatitis C Therapy–predicting Effects of Treatment. Nature reviews Gastroenterology & hepatology. 2015 aug; 12(8):437–45. doi: 10.1038/nrgastro.2015.97

6.   Boianelli A, Sharma-Chawla N, Bruder D, Hernandez-Vargas EA. Oseltamivir PK/PD modelling and Simulation to Evaluate Treatment Strategies Against Influenza-Pneumococcus Coinfection. Frontiers in Cellular and Infection Microbiology. 2016; 6(60). doi: 10.3389/fcimb.2016.00060 PMID: 27379214

7.   Altrock PM, Liu LL, Michor F. The Mathematics of Cancer: Integrating Quantitative Models. Nature Reviews Cancer. 2015 nov; 15(12):730–745. doi: 10.1038/nrc4029 PMID: 26597528

8.   Le Novère N. Quantitative and Logic Modelling of Molecular And Gene Networks. Nature reviews Genetics. 2015 feb; 16(3):146–58. doi: 10.1038/nrg3885 PMID: 25645874

9.   Miao H, Xia X, Perelson AS, Wu H. On Identifiability of Nonlinear ODE Models and Applications in Viral Dynamics. SIAM review Society for Industrial and Applied Mathematics. 2011 Jan; 53(1):3–39. doi: 10.1137/090757009 PMID: 21785515

10. Duvigneau S, Sharma-Chawla N, Boianelli A, Stegemann-Koniszewski S, Nguyen VK, Bruder D, et al. Hierarchical Effects of Pro-inflammatory Cytokines on The Post-influenza Susceptibility to Pneumococcal Coinfection. Scientific Reports. 2016 Nov;. doi: 10.1038/srep37045 PMID: 27872472

11. Hernandez-Vargas EA, Wilk E, Canini L, Toapanta FR, Binder SC, Uvarovskii A, et al. Effects of Aging on Influenza Virus Infection Dynamics. Journal of Virology. 2014 Apr; 88(8):4123–31. doi: 10.1128/JVI.03644-13 PMID: 24478442

12. Yates A, Stark J, Klein N, Antia R, Callard R. Understanding the Slow Depletion of Memory CD4+ T Cells in HIV Infection. PLoS medicine. 2007 may; 4(5):e177. doi: 10.1371/journal.pmed.0040177 PMID: 17518516

13. Raue a, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. Bioinformatics (Oxford, England). 2009 Aug; 25(15):1923–9. doi: 10.1093/bioinformatics/btp358

14. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally Sloppy Parameter Sensitivities in Systems Biology Models. PLoS Comput Biol. 2007 Oct; 3(10):1871–1878. doi: 10.1371/journal.pcbi.0030189 PMID: 17922568

15. Wu H, Zhu H, Miao H, Perelson AS. Parameter Identifiability and Estimation of HIV/AIDS Dynamic Models. Bull Math Biol. 2008; 70(3):785–799. doi: 10.1007/s11538-007-9279-9 PMID: 18247096

16. Li P, Vu QD. Identification of Parameter Correlations for Parameter Estimation in Dynamic Biological Models. BMC Syst Biol. 2013; 7:91. doi: 10.1186/1752-0509-7-91 PMID: 24053643

17. Chis OT, Banga JR, Balsa-Canto E. Structural Identifiability of Systems Biology Models: a Critical Comparison of Methods. PloS one. 2011 Jan; 6(11):e27755. doi: 10.1371/journal.pone.0027755 PMID: 22132135

18. Brun R, Reichert P, Kuensch HR. Practical Identifiability Analysis of Large Environmental Simulation Models. Water Resources Research. 2001; 37(4):1015–1030. doi: 10.1029/2000WR900350

19. Xia X, Moog CH. Identifiability of Nonlinear Systems with Application to HIV/AIDS Models. IEEE Transactions on Automatic Control. 2003 Feb;(2):330–336. doi: 10.1109/TAC.2002.808494

20. Cedersund G. In: Geris L, Gomez-Cabrero D, editors. Prediction Uncertainty Estimation Despite Unidentifiability: an Overview of Recent Developments. Cham: Springer International Publishing; 2016. p. 449–466.

21. Smith AM, Perelson AS. Influenza A Virus Infection Kinetics: Quantitative Data and Models. Syst Biolo Med. 2011; 3(4):429–45.

22. Pawelek KA, Huynh GT, Quinlivan M, Cullinane A, Rong L, Perelson AS. Modeling Within-host Dynamics of Influenza Virus Infection Including Immune Responses. PLoS Comput Biol. 2012; 8(6):e1002588. doi: 10.1371/journal.pcbi.1002588 PMID: 22761567

23. Petrie SM, Guarnaccia T, Laurie KL, Hurt AC, McVernon J, McCaw JM. Reducing Uncertainty in Within-Host Parameter Estimates of Influenza Infection by Measuring Both Infectious and Total Viral Load. PLoS ONE. 2013; 8(5):e64098. doi: 10.1371/journal.pone.0064098 PMID: 23691157

24. Boianelli A, Pettini E, Prota G, Medaglini D, Vicino A. Identification of a Branching Process Model for Adaptive Immune Response. In: 52nd IEEE Conference on Decision and Control; 2013. p. 7205–7210.

25. Pubmed Central. PMC Policies; 2016.

26. Nguyen VK, Hernandez-Vargas EA; International Federation of Automatic Control. Identifiability Challenges in Mathematical Models of Viral Infectious Diseases. 17th IFAC Symposium on System Identification. 2015 Oct;.

27. Baccam P, Beauchemin C, Macken Ca, Hayden FG, Perelson AS. Kinetics of Influenza A Virus Infection in Humans. Journal of Virology. 2006 Aug; 80(15):7590–9. doi: 10.1128/JVI.01623-05 PMID: 16840338

28. Moehler L, Flockerzi D, Sann H, Reichl U. Mathematical Model of Influenza A Virus Production in Large-scale Microcarrier Culture. Biotechnology and Bioengineering. 2005; 90(1):46–58. doi: 10.1002/bit.20363

29. Handel A, Longini IM, Antia R. Towards a Quantitative Understanding of the Within-host Dynamics of Influenza A Infections. Journal of the Royal Society, Interface / the Royal Society. 2010 Jan; 7(42):35–47. doi: 10.1098/rsif.2009.0067

30. Holder BP, Simon P, Liao LE, Abed Y, Bouhy X, Beauchemin CAA, et al. Assessing the in vitro Fitness of an Oseltamivir-resistant Seasonal A/H1N1 Influenza Strain Using a Mathematical Model. PloS one. 2011 Jan; 6(3):e14767. doi: 10.1371/journal.pone.0014767 PMID: 21455300

31. Wu H, Ding AA, De Gruttola V. Estimation of HIV Dynamic Parameters. Statistics in Medicine. 1998; 17:2463–2485. doi: 10.1002/(SICI)1097-0258(19981115)17:21%3C2463::AID-SIM939%3E3.0.CO;2-A PMID: 9819839

32. Kirschner D. Using Mathematics to Understand HIV Immune Dynamics. AMS notices. 1996; 43 (2):191–202.

33. Hernandez-Vargas EA, Middleton RH. Modeling the Three Stages in HIV Infection. Journal of Theoretical Biology. 2013; 320:33–40. doi: 10.1016/j.jtbi.2012.11.028 PMID: 23238280

34. Duffin RP, Tullis RH. Mathematical Models of the Complete Course of HIV Infection and AIDS. Journal of Theoretical Medicine. 2002; 4(4):215–221. doi: 10.1080/1027366021000051772

35. Nguyen VK, Binder SC, Boianelli A, Meyer-Hermann M, Hernandez-Vargas EA. Ebola Virus Infection Modelling and Identifiability Problems. Frontiers in Microbiology. 2015; 6(257).

36. Rawlins EL, Hogan BLM. Ciliated Epithelial Cell Lifespan in the Mouse Trachea and Lung. American journal of physiology Lung cellular and molecular physiology. 2008 July; 295(1):L231–4. doi: 10.1152/ajplung.90209.2008 PMID: 18487354

37. Pinilla LT, Holder BP, Abed Y, Boivin G, Beauchemin C. The H275Y Neuraminidase Mutation of the Pandemic A/H1N1 Influenza Virus Lengthens The Eclipse Phase And Reduces Viral Output of Infected Cells, Potentially Compromising Fitness in Ferrets. Journal of virology. 2012 Oct; 86(19):10651–60. doi: 10.1128/JVI.07244-11 PMID: 22837199

38. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics. Springer; 2009.

39. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis; 2013.

40. Ma S, Kosorok MR. Robust Semiparametric M-estimation and the Weighted Bootstrap. Journal of Multivariate Analysis. 2005; 96(1):190–217. doi: 10.1016/j.jmva.2004.09.008

41. Efron B, Hinkley DV. Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information. Biometrika. 1978; 65(3):457–483. doi: 10.1093/biomet/65.3.457

42. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014.

43. Mullen K, Ardia D, Gil D, Windover D, Cline J. DEoptim: An R Package for Global Optimization by Differential Evolution. Journal of Statistical Software. 2011; 40(6):1–26. doi: 10.18637/jss.v040.i06

44. Byrd RH, Lu P, Nocedal J, Zhu C. A Limited Memory Algorithm for Bound Constrained Optimization. SIAM J Sci Comput. 1995 Sep; 16(5):1190–1208. doi: 10.1137/0916069

45. Thiébaut R, Guedj J, Jacqmin-Gadda H, Chêne G, Trimoulet P, Neau D, et al. Estimation of Dynamical Model Parameters taking into account Undetectable Marker Values. BMC medical research methodology. 2006 Jan; 6(38):1–9.

46. Simonoff JS. Smoothing Methods in Statistics. Springer Series in Statistics. Springer; 1996.

47. Wood SN. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2011; 73(1):3–36. doi: 10.1111/j.1467-9868.2010.00749.x

48. Smith AM, Adler FR, McAuley JL, Gutenkunst RN, Ribeiro RM, McCullers JA, et al. Effect of 1918 PB1-F2 Expression on Influenza A Virus Infection Kinetics. PLoS Comput Biol. 2011 Feb; 7(2):e1001081. doi: 10.1371/journal.pcbi.1001081 PMID: 21379324

49. Berkson J. Are There Two Regressions? Journal of the American Statistical Association. 1950; 45 (250):pp. 164–180. doi: 10.1080/01621459.1950.10483349

50. Althubaiti A, Donev A. Non-Gaussian Berkson Errors in Bioassay. Statistical Methods in Medical Research. 2012;. PMID: 23070587

51. Toapanta FR, Ross TM. Impaired iMmune Responses in the Lungs of Aged Mice Following Influenza Infection. Respir Res. 2009; 10:112. doi: 10.1186/1465-9921-10-112 PMID: 19922665

52. Heldt FS, Frensing T, Pflugmacher A, Gropler R, Peschel B, Reichl U. Multiscale Modeling of Influenza A Virus Infection Supports the Development of Direct-acting Antivirals. PLoS Comput Biol. 2013; 9(11): e1003372. doi: 10.1371/journal.pcbi.1003372 PMID: 24278009

53. Dobrovolny HM, Reddy MB, Kamal MA, Rayner CR, Beauchemin CA. Assessing Mathematical Models of Influenza Infections using Features of the Immune Response. PLoS ONE. 2013; 8(2):e57088. doi: 10.1371/journal.pone.0057088 PMID: 23468916

54. Ashyraliyev M, Fomekong-nanfack Y, Kaandorp JA, Blom JG. Systems Biology: Parameter Estimation for Biochemical Models. The FEBS Journal. 2009; 276:886–902. doi: 10.1111/j.1742-4658.2008. 06844.x PMID: 19215296

55. Haario H, Laine M, Mira A, Saksman E. DRAM: Efficient Adaptive MCMC. Statistics and Computing. 2006; 16(4):339–354. doi: 10.1007/s11222-006-9438-0

56. Kesar Singh MX. Bootlier-Plot: Bootstrap Based Outlier Detection Plot. Sankhyā: The Indian Journal of Statistics (2003–2007). 2003; 65(3):532–559.

57. Huber PJ, Ronchetti EM. Robust Statistics. Wiley Series in Probability and Statistics. Wiley; 2009.

58. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index. Computer Physics Communications. 2010; 181(2):259–270. doi: 10.1016/j.cpc.2009.09.018

59. Sobol IM. Global Sensitivity Indices for Nonlinear Mathematical Models and their Monte Carlo Estimates. Mathematics and Computers in Simulation. 2001; 55(1–3):271–280. The Second IMACS Seminar on Monte Carlo Methods. doi: 10.1016/S0378-4754(00)00270-6

60. Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2011; 73(2):123–214. doi: 10.1111/j. 1467-9868.2010.00765.x

61. Ramsay JO, Hooker G, Campbell D, Cao J. Parameter Estimation for differEntial Equations: a Generalized Smoothing Approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69(5):741–796. doi: 10.1111/j.1467-9868.2007.00610.x

62. Liang H, Wu H. Parameter Estimation for Differential Equation Models Using a Framework of Measurement Error in Regression Models. J Am Stat Assoc. 2008 Dec; 103(484):1570–1583. doi: 10.1198/ 016214508000000797 PMID: 19956350

63. Hagen DR, White JK, Tidor B. Convergence in Parameters and Predictions using Computational Experimental Design. Interface Focus. 2013 Aug; 3(4):20130008. doi: 10.1098/rsfs.2013.0008 PMID: 24511374

64. Raue, A, Kreutz, C, Theis, FJ, Timmer, J. Joining Forces of Bayesian and Frequentist Methodology: A Study for Inference in the Presence of Non-Identifiability. ArXiv e-prints. 2012 Feb;.