AMERICAN SOCIETY FOR MICROBIOLOGY
Journals.ASM.org

# Design, optimization, and inference of multiphasic decay of infectious virus particles

**Jérémy Seurat[1], Krista R. Gerbino[2], Justin R. Meyer[2], Joshua M. Borin[2], Joshua S. Weitz[3,4,1,*]**

[1]Institut de Biologie, École Normale Supérieure, 75005 Paris, France.

[2]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA.

[3]Department of Biology, University of Maryland, College Park, MD 20742, USA.

[4]Department of Physics, University of Maryland, College Park, MD 20742, USA

[*]Address correspondence to Joshua S. Weitz, jsweitz@umd.edu

## ABSTRACT

The loss of virus particles is typically considered to arise from a first-order kinetic process. Signals of deviations from this exponential decay are often de-prioritized. Here, we propose methods to evaluate if a design is adequate to evaluate evidence for multiphasic virus particle decay and to optimize the sampling times of decay experiments, accounting for uncertainties in viral kinetics. First, we evaluate 1500 synthetic scenarios of biphasic decays, with varying decay rates and initial proportions of subpopulations. Robust inference of multiphasic decay is more likely when the faster decaying subpopulation predominates insofar as early samples are taken to resolve the faster decay rate. Overall, we find that design optimization leads to a better precision of estimation while reducing the number of samples. It helps to estimate adequately the fastest decay in

54% of situations vs. 41% using a non-optimized design. We then apply these methods to infer multiple decay rates associated with the decay of ΦD9, an evolved isolate derived from phage Φ21. A pilot experiment confirmed that ΦD9 decay is multiphasic, but was unable to resolve the rate or proportion of the fast decay subpopulation(s). We then applied optimal design methods to propose new ΦD9 sampling times. Using this strategy, we were able to robustly estimate both decay rates and their respective subpopulations. Notably, we conclude that the vast majority (94%) of the population decays at a rate 16-fold higher than a slow decaying population. Altogether, these results provide methods to quantitatively estimate heterogeneity in viral decay.

**keywords**: multiphasic decay, viral decay, optimal design, inference, Fisher information matrix

## INTRODUCTION

The life cycle of viruses includes a balance between production and decay. Although production of virus particles from infected hosts is often the priority of environmental and health-associated studies, the numerical abundance of viruses in an environment is also shaped by the rate at which virus particles (i.e., virions) lose their infectivity over time. Mechanistically, virions can lose infectivity due to multiple processes including thermal-induced destabilization, aggregation, physical decay (e.g., the separation of tail fibers from the head), and/or damage to the viral genome (e.g., via exposure to UV light). The aggregate impact of these processes is conventionally modeled as a single, kinetic rate [6]. The decay rate can be considered as the rate constant associated with the exponential decay of infectious virions over time and is often strongly linked to temperature [6, 3]. However, virus populations need not always decay exponentially [4, 7, 21].

A virus decay curve is a monotonically decreasing function over time, and the simplest exception to that of pure exponential decay is termed 'multiphasic decay'. A population exhibiting multiphasic decay would be characterized by a set of different exponential rate constants. Hence, each subpopulation exhibits pure exponential decay but the total population does not. Multiphasic decay could arise because the population is

comprised of different viral 'types' (e.g., species or alternatively self-assembled particle morphs) or because the process of decay involves different routes to failure. Irrespective of the underlying mechanism, inferring multiphasic decay presents unique challenges to experimental design. In general, in the case of biphasic decay, a biexponential model is used to describe the data by estimating the two slopes corresponding to each decay rate. As such, conventional approaches to equitemporal spacing of measurements may fail to resolve the differences between rate constants and the magnitude of the subpopulation corresponding to each of the different decay rates.

The design choice for viral decay experiments is therefore crucial. In this context, the study design includes the number of measurements and designation of the time of each measurement. Frequently, the design is chosen in an empirical way, for instance with equi-spaced sampling times. A wrong design choice results in poor parameter precision. For instance, the first and fastest slope is wrongly estimated if early time samples are not collected. Also, in the case of short studies, the slowest slope, corresponding to the more stable subpopulation, cannot be observed if there has not been enough measurements of decay when this is the only population left. Instead, to approach the inference problem for multiphasic decay requires evaluating and optimizing the design elements, i.e. the duration time of the study as well as the number and the allocation of the measurement times.

The problem of design choice can be solved by choosing the experimental design appropriately. Optimal design [1, 18] has shown good performance in decreasing the bias and uncertainty in parameter estimation in multiple contexts, including in virus studies [13, 20]. However, experimentalists often have limited prior information on parameters associated with viral decay rates and the relative proportion of each decay type in a population. In this case, robust design optimization accounting for these uncertainties may be required [12]. In this context, the design objective is to select the number of replicates, the number of samples, the sampling times, and (potentially) the initial viral density.

Here, we use a Fisher information matrix-based design optimization method to precisely infer rates and magnitudes of subpopulations for multiphasic viral decay experiments. This enables us to investigate the importance of the viral parameters (decay rates, initial proportions of each subpopulation) when choosing the design and to show how it optimizes the design, even in the case of parameter uncertainties. By leveraging *in silico* studies, we demonstrate a proof of principle for robust design and then practically apply these methods to the phage ΦD9, showing that we are able to successfully recapitulate multiphasic decay. This inference approach identified a hidden majority of fast decaying phage and a subpopulation of slower decaying phage, raising new questions on the mechanisms and eco-evolutionary dynamics associated with variation in phage loss rates.

# MATERIALS AND METHODS

## Optimal design for time-series experiments

### *Statistical model*

A nonlinear model to describe viral density observations $y$ (of length $n$) could be written as follows (eq. 1):

$$y = f(\theta, \xi) + \varepsilon \tag{1}$$

The function $f$ (the structural model), depending on $\theta$, the vector of the parameters describing the viral kinetics, and $\xi$, the design, provides a $n$-vector of model predicted values. The $n$-vector of errors is denoted $\varepsilon$, with $\varepsilon \sim \mathcal{N}(0, (\sigma \times f(\theta, \xi))^2)$, i.e. assuming a proportional error to the structural model $f$, with $\sigma$ the standard deviation of proportional error terms. The vector of parameters, of size $P$, to be estimated is composed of $\theta$ and $\sigma$.

### *Fisher information matrix based design evaluation and optimization*

A design $\xi$ is defined by the number of samples $n$ and their allocation in time ($t_1$,...,$t_n$), and some additional design variables (e.g. the initial viral density $V_0$ which can otherwise be defined as a parameter).

Optimizing the design in our case is optimizing the choice of samples (number, time allocation) to maximize precision of parameter estimation. According to the Rao-Cramer inequality, the expected Fisher information matrix (FIM or $M_F(\theta, \xi)$) is the inverse of the lower bound of the variance-covariance matrix of any unbiased estimated parameters. More details about FIM evaluation for nonlinear models are given in [1, 11, 23]. Therefore, the square root of the inverse of diagonal elements of the FIM is used to provide expected standard errors (SE) of parameters.

The D-optimality criterion $\Phi_D$ (eq. 2), widely used in the field of optimal design, consists in the determinant of the FIM normalized by $P$ (the number of model parameters to be estimated) [1]:

$$\Phi_D = |M_F(\theta, \xi)|^{(1/P)} \tag{2}$$

Optimization algorithms aim to find the design (given the constraints) which maximizes this criterion, i.e. which gives the best overall expected precision of estimated parameters.

The D-optimality criterion is used to perform local design evaluation or optimization, i.e. considering the model and parameter as known. Alternatively, in case of parameter uncertainty, robust design optimization can be performed. Several methods and criteria exist for this purpose, and the method based on the HClnD-optimality criterion $\Phi_{HClnD}$ (eq. 3 is efficient [12], assuming an expected parameter distribution:

$$\Phi_{HClnD} = \sum_{i=1}^{m} \frac{1}{mq} \ln |M_F(\theta, \xi)| \tag{3}$$

where there are $q$ uncertain parameters and $m = 2^q$. In this case $\theta$ are lower or upper percentile values of parameter distribution. In our study, we investigate robust designs according to this optimality criterion assuming 1 uncertain parameter (1UP or Robust 1UP) or 3 uncertain parameters (3UP or Robust 3UP).

In this study, optimizers are used to find the best allocation of sampling times, given $n$. The first category of algorithm performs optimization in a discrete design space. Discrete optimization requires defining the vector of all the possible sampling time allocations. We used the Fedorov-Wynn discrete algorithm [10, 27] implemented in PFIM version 4 [8] to find the optimal combination of sampling times for the illustrative studies of Mu/P1 and M13/P2, as well as for the phage ΦD9 decay experiment. The other category consists in algorithms working in a continuous design space, for which the sampling time window(s) and initial values for the design should be defined. We used the simplex continuous algorithm [19] implemented in PFIM version 4 to find the optimal sampling times according to viral parameters and the influence of design constraints (maximal duration and number of samples).

### *Performance evaluation*

The designs are compared in function of their optimality criteria $\Phi_X$ defined above. The X-efficiency $E_X$ of a design $\xi$, with respect to a reference design $\xi_R$ (for instance a non-optimized design), is computed as $E_X(\%) = \frac{\Phi_X(\xi)}{\Phi_X(\xi_R)} \times 100$. The expected standard errors of parameters are the main element of interest when evaluating a design. Relative standard errors (RSE, given in percentages, eq. 4) thresholds can be defined.

$$RSE(\theta) = \frac{SE(\theta)}{\theta} \times 100 \tag{4}$$

In this work, we consider that good RSE are below 30%, whereas poor and very poor RSE are greater or equal than 50% and 100%, respectively.

## Application to ΦD9 multiphasic decay

### *Viral decay experiment*

For each replicate a separate phage stock was created by coculturing 2µL of frozen isogenic phage ΦD9 stock preserved at 80°C with $5 \times 10^8$ ΔLamB ΔOmpC *E. coli* cells from an overnight culture into 10mL TrisLB (see Borin et al. [4] for media recipe and strain

descriptions) at 37°C, 120rpm for 4 hours. The double receptor knockout host was used to select for the maintenance of OmpF use, which was previously found to be evolutionarily unstable. The relatively short incubation time was chosen to reduce the loss of fast decaying phage particles. Twelve phage lysates were filtered through 0.22 µm membranes and then a subsample of each stock was immediately plated to estimate initial viral titers. Plating was done on soft agar overlays with *E. coli* strain BW25113 and incubated overnight at 37°C [4]. Plaques were then enumerated. After initial phage samples were taken, 4mL of each lysate was then transferred to glass tubes that phages are not known to bind to [24] and incubated at 37°C and stationary to reduce evaporation. Additional phage samples were plated at 1, 2, 3, 5, 8, 24, 48, and 72 hours after the initial filtering.

### *Data fitting and design optimization*

Using the estimated parameters after data fitting from this experiment, robust design optimization were performed according to HClnD-criterion (3), supposing a possible triphasic decay, and assuming uncertain parameters for the two first phases and known parameters for the last phase and the initial viral density. The following design constraints were defined: a total of 9 sampling times including 0, 1, 2 and 3 days, and optimizing the 5 other sampling times among the vector (0.04, 0.08, 0.12, 0.17, 0.21, 0.25, 0.29, 0.33, 0.67, 0.71, 0.75, 0.79, 0.83, 0.88, 0.92, 0.96) d.

Population parameters were estimated by fitting the combined data from the replicates of an experiment, considering between-replicates differences as noise (the "naive pooled approach") and following a maximum likelihood based method. Models were compared between them using the Akaike information criterion (AIC), which accounts for likelihood and the number of estimated parameters.

### **Software**

R was used to perform design evaluation and optimization, through the PFIM version 4.0. program [8] and figures. Data fitting was performed with Monolix version 2021R2 [16]. Codes for design optimization and data fitting are available in Supplementary files.

## Data availability

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The codes to perform design evaluation and optimization, as well as data fitting are accessible at: https://github.com/jeremyseurat/DesignMultiphasicViralDecay.

# RESULTS

## Impact of multiphasic rates of population decay

The decay of virus populations is impacted by the number of 'phases', corresponding decay constants, and the proportions of each subpopulation in the population as a whole. To explore these effects, we illustrate the impact of biphasic decay on total virus populations when modulating the associated decay rate and proportion of subpopulations. In case of a monophasic decay, the viral load ($V$) decayse over time ($t$) as follows:

$$V(t) = V_0 \times \exp\left(-\alpha \times t\right) \tag{5}$$

with $V_0$ the initial viral density which could be estimated or known from the study settings and $\alpha$ the decay rate. In this case, as long as viral density is quantifiable, a limited number of measurements (e.g. two or three) over time are sufficient to estimate $\alpha$ with adequate precision. In case of biphasic decay, the viral load decays according to the following:

$$V(t) = V_0 \times \left(f_a \times \exp\left(-\alpha \times t\right) + f_b \times \exp\left(-\beta \times t\right)\right) \tag{6}$$

with the parameters $f_a$ the initial proportion of the phage of decay rate $\alpha$ and $f_b$ for the phage of decay rate $\beta$, $f_a = 1 - f_b$. The impact of the different parameters (decay rates and initial proportions) on the viral kinetics is shown in FIG **1**. The two decay phases can be conveniently distinguished, in case of two very different decay rates, for example if $\alpha = 0.5$ /d and $\beta = 0.05$ /d with similar proportions of the two subpopulations. The two slopes could also be readily observable if the proportion of the subpopulation decreasing the fastest is greater than that decreasing the slowest (e.g. $\alpha = 0.1$ /d, $\beta = 0.5$ /d,

$f_a = 0.05$ and $f_b = 0.95$). However, distinguishing the two phases is more complex when decay rate values are close to each other or when the slowest-declining viral subpopulation represents a large proportion of the total.

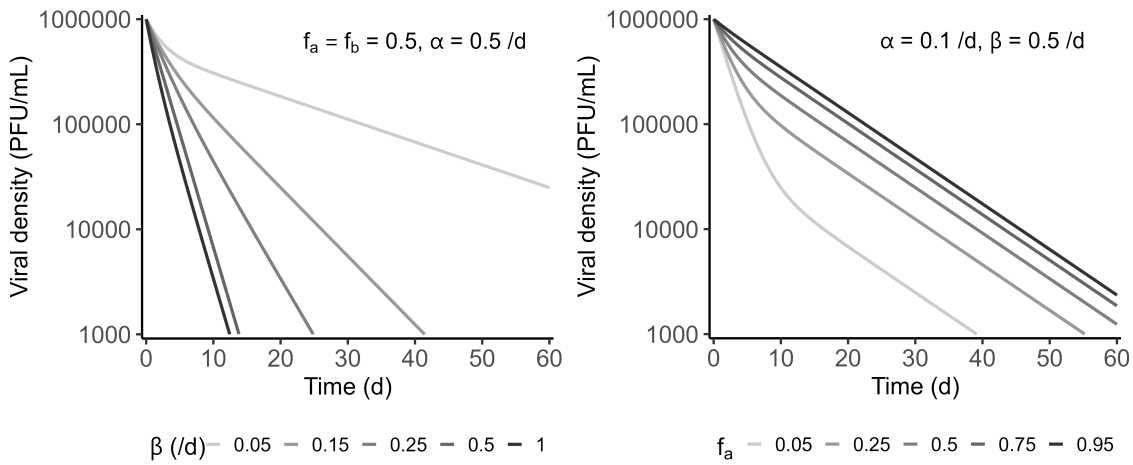



**FIG 1 Influence of the viral parameters on a biphasic decay.** Biphasic decays (eq. 6) are represented on the two panels: with different decay rates for one population ($\beta$) ranging from 0.05 to 1 /d (with same initial proportions for both populations i.e. $f_a$=$f_b$=0.5 and decay rate of the other population ($\alpha$) of 0.35 /d) on the left and different initial proportions $f_a$ ranging from 0.05 to 0.95 (with $f_b$=1-$f_a$, $\alpha$=0.1 /d and $\beta$=0.3/d) on the right.

## *Decay rates, inference and design optimization: illustrative example*

To illustrate how the viral decay rates, and the design (measurement times) choice can influence the precision of parameter estimation, we choose two different examples composed of two phage populations with known decay rates [6].

In the first example, we consider a study composed of two phages with very different decay rates such as Mu (with a decay rate of 0.29 /d) and P1 (with a decay rate of 0.077 /d). We assume similar proportions of the two phages ($f_a$=$f_b$=0.5), and an error parameter (notably reflecting the noise) $\sigma$ of 0.2 (i.e. standard deviation of the discrepancy between data and real viral density value of 20% of the real value). Using a non-optimized design with 7 equi-spaced sampling times (1 sample every 10 days), the precision of parameters

relative to the fastest phage, Mu are poor with a relative standard error (RSE) of 101% for the estimation of the decay rate of Mu, which is due to the lack of information in the early kinetic.

For the same viruses, if we fix the initial sampling time (t=0 d) and optimize the 6 other sampling times (among (2, 4, 6 ,8, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 35, 40, 50, 60) d), the optimal design is composed of the following sampling times: (0, 4, 6, 20, 22, 50, 60) d. This design allows a better overall precision of parameters (gain of D-efficiency of 24%) and the RSE of the decay rate of Mu is expected to be 63% (vs. 101% without design optimization, TABLE 1). However, if the two phages have a similar decay rate, for instance M13 (0.074 per day) and P2 (0.041 per day), the expected precision of the different parameters is very poor (RSE>100%) even using an optimized design.

**TABLE 1 Viral decay rates effect on expected precision of estimates from a non-optimized or optimized design.** The phage subpopulations Mu and P1 are represented in the top part and M13 and P2 in the bottom part. The vector of sampling times is given in days (d), $\alpha$ is the decay rate of the slowest virus (Mu and M13, respectively), $\beta$ is the decay rate of the fastest virus (P1 and P2, respectively), $f_a$ and $f_b$ are the initial proportions (corresponding to $\alpha$ and $\beta$, respectively), $\sigma$ is the proportional error parameter, RSE is the relative standard error: a RSE greater than 50% or 100% is indicated by a light grey or a dark grey cell, respectively.

| | | **Non-optimized** | **Optimized** |
|---|---|---|---|
| **Mu/P1** | | **(0, 10, 20, 30, 40, 50, 60) d** | **(0, 4, 6, 20, 22, 50, 60) d** |
| **Parameter (unit)** | **Value** | **RSE (%)** | **RSE (%)** |
| $\alpha$ (/d) | .077 | 10 | 8 |
| $\beta$ (/d) | .29 | 101 | 63 |
| $f_a$ | .50 | 34 | 29 |
| $f_b$ | .50 | 52 | 46 |
| $\sigma$ | .20 | 27 | 27 |
| **M13/P2** | | **(0, 10, 20, 30, 40, 50, 60) d** | **(0, 2, 14, 16, 35, 40, 60) d** |
| **Parameter (unit)** | **Value** | **RSE (%)** | **RSE (%)** |
| $\alpha$ (/d) | .041 | 225 | 214 |
| $\beta$ (/d) | .074 | 335 | 317 |
| $f_a$ | .50 | 812 | 771 |
| $f_b$ | .50 | 799 | 317 |
| $\sigma$ | .20 | 27 | 27 |

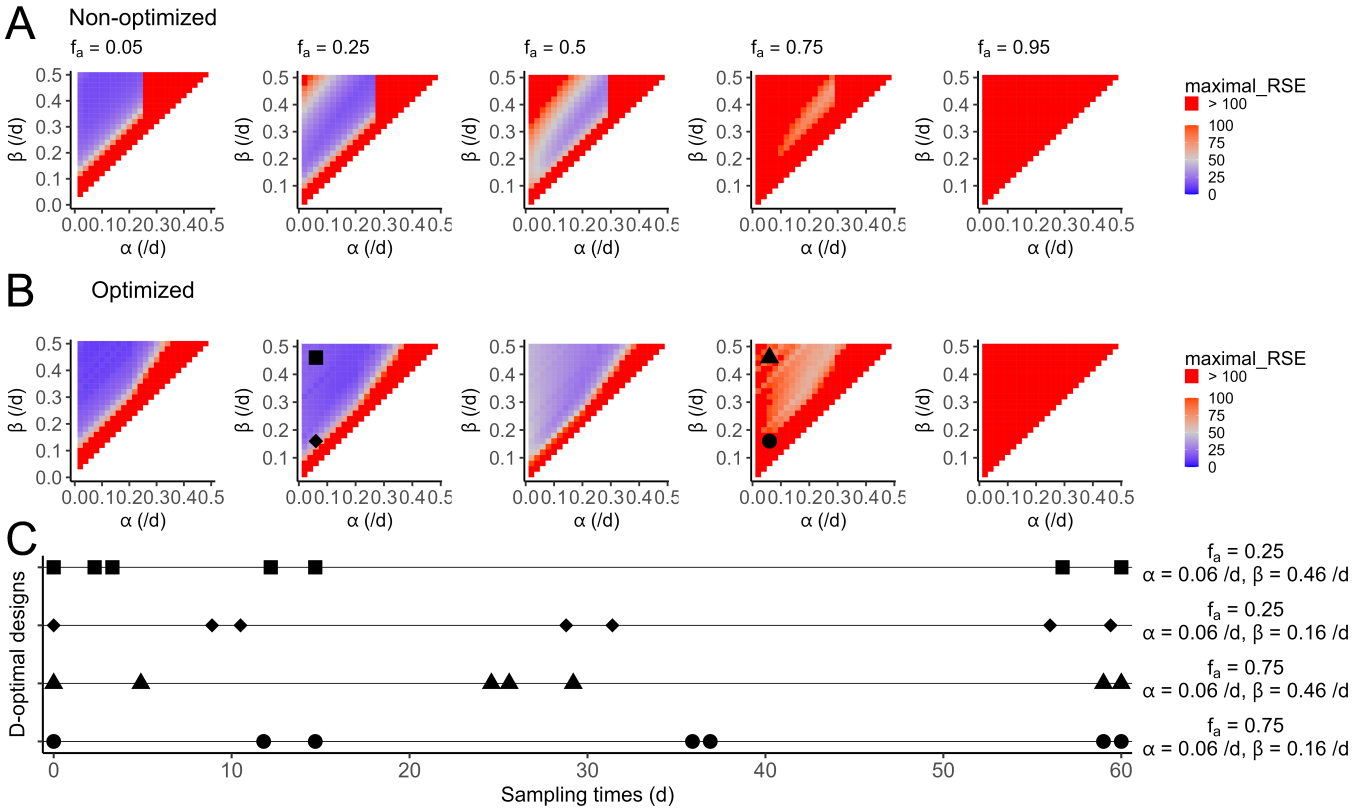
## *Optimal time measurements according to viral parameters*

Next, we studied more extensively the influence of the decay rate of the two viral subpopulations on the RSE of resulting fits (different combinations of decay rates between 0.02 and 0.5 /d, FIG 2). Assuming equal initial proportions of the two viruses, a non-optimized design (1 sampling time every 10 days during 60 days) leads to poor RSE

and potentially the non-identifiability of parameters if $\beta$ (the fastest decay rate) is higher than 0.35 /d and $\alpha$ (the slowest) is lower than 0.15 /d or higher than 0.3 /d. An optimal design composed of the same number of sampling times found using the simplex algorithm (i.e. finding the 7 optimal times in the continuous design space [0,60] d) can provide acceptable RSE in most of those days. However, as in the example shown in TABLE 1, two very close decay rates for instance which differ from 0.06 /d are always difficult to estimate properly, even with design optimization.

From the influence of the initial proportion of the two viral subpopulations $f_a$ and $f_b$ (FIG 1), we understand that the identification of the two phases of the decay is easier with a low proportion of viruses that have the slower decay rate. This is confirmed by FIG 2 where the maximal RSE are never below 100% if $f_a$=0.95 regardless of $\alpha$ and $\beta$ values between 0.02 and 0.5 /d, even if the 7 measuring times are optimized in the [0,60] days window.

However, design optimization helps to have a better precision of estimation when the initial proportions are more balanced ($f_a$ between 0.25 and 0.75), especially when one of the decay rate is very fast compared to the other one. Interestingly, the optimization algorithm tends to find optimal measurement times earlier when one of the two decay rates is high than when both decay phases are slow (e.g. $\alpha$ = 0.06 /d, $\beta$ = 0.46 /d vs. $\alpha$ = 0.06 /d, $\beta$ = 0.16 /d, FIG 2). In case of very low initial proportion of the slowest-decaying virus ($f_a$=0.05), a non-optimized design is generally good enough to provide sufficient RSE in most of the cases.




258

**FIG 2 Influence of viral parameters on maximal imprecision (RSE) of estimates using an (A) empirical non-optimized design or an (B) optimized design and on (C) optimal sampling times.** The non-optimized design consists in 7 measuring times from 0 to 60 days every 10 days. The design optimization is performed by the simplex algorithm (7 times in the [0-60] days window), where $\alpha$ is the decay rate of the slowest virus, $\beta$ is the decay rate of the fastest virus, $f_a$ and $f_b$=1-$f_a$ are the initial proportions (corresponding to $\alpha$ and $\beta$, respectively). RSE refers to relative standard error and the color represents the higher expected RSE among the estimated parameters: $\alpha$, $\beta$, $f_a$, $f_b$ and $\sigma$ (the error model parameter). Optimal designs are represented for different parameters combinations by specific symbols ($\alpha$ = 0.06 /d, square: $\beta$=0.46/d and $f_a$=0.25, rotated square: $\beta$=0.16/d and $f_a$=0.25, triangle: $\beta$=0.46/d and $f_a$=0.75, circle: $\beta$=0.16/d and $f_a$=0.75).

## *Influence of design constraints*

Another possibility to reach sufficient information from the design is to increase the number of sampling times during the study. In this section, we evaluate how many measurements would be needed to achieve precise parameter estimates (RSE < 30%), depending on the decay rates of the two populations which are assumed to be in the same initial proportions (i.e. $f_a = f_b = 0.5$). Moreover, we investigate if considering longer studies (120 days instead of 60 days) can help to quantify precisely the two slopes during biphasic decay in this particular case study.

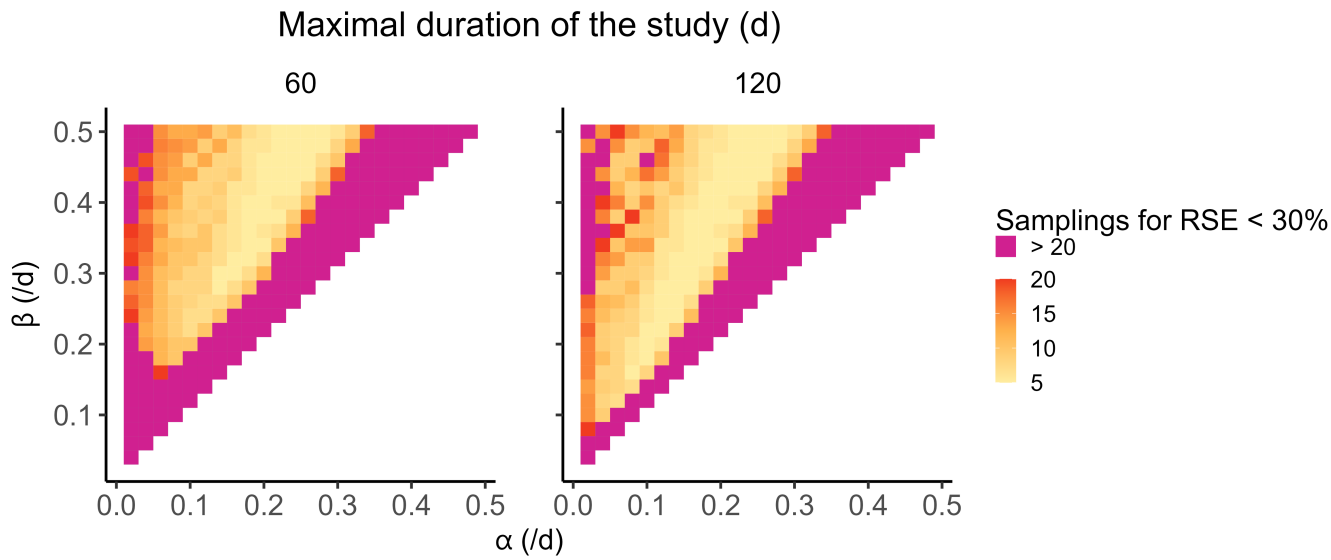



**FIG 3 Influence of maximal duration time on needed sampling times to reach adequate precision of estimation.** A flexible number of measuring times is considered between 0 and 60 days on the left and between 0 and 120 days on the right, respectively. Sampling times of the design are optimized using the simplex algorithm in the corresponding sampling window ([0,60] and [0,120]). $\alpha$ is the decay rate of the slowest virus, $\beta$ is the decay rate of the fastest virus, the initial proportion of each viral population is 0.5. RSE: Relative Standard Error.

Increasing the number of measurements increases the quantity of information, and as a consequence improves the precision of parameters. For instance, in the case of $\alpha = 0.1$ /d and $\beta = 0.4$ /d, 7 optimized samplings during the 60 days following the beginning of the study are not sufficient to obtain RSE < 30% in all parameters. In this specific case, 10

sampling times are needed to reach this objective (FIG 3). This option however increases the cost of the study. Furthermore, optimizing sampling times may already provide sufficient information. In our example, there are cases where 5 measuring times would be sufficient to obtain precise estimates, for example if $\alpha$ = 0.24 /d and $\beta$ = 0.44 /d.

The duration of the study is another important design consideration which depends on the two decay rates. For example when $\alpha$ is 0.1 /d and $\beta$ is 0.2 /d, allowing a maximal duration of 120 days instead of 60 days can reduce the number of measurements by a factor 2 (5 vs. 10, respectively) to reach the RSE objective. Importantly, extending the duration of the study is needed to investigate the decay of two viruses with slow elimination: for instance if $\alpha$ is 0.08 /d and $\beta$ = 0.12/d, choosing a maximal duration of 120 days enables precise parameter estimation, which is not the case with a 60-day study design, regardless of the number or the allocation of the sampling times. In constrast, the design can be shortened if we know that one of the two populations have a fast decay rate. Indeed, similar number of sampling times are needed between 60 and 120 days when $\beta$ is between 0.4 and 0.5 /d (FIG 3).

## Finding optimal design accounting for parameter uncertainty

In many cases, we do not know the parameters describing the viral density over time. A possibility is to assume the uncertainty on one or several parameters, and define a prior distribution. Robust designs account for the parameter uncertainty during the optimization procedure. Defining priors based on the distribution of viral decays in [6], we investigated two robust design scenarios: accounting for uncertainty in i) one viral decay rate, i.e. 1 Uncertain Parameter (1UP), the other one set at 0.09 /d and initial proportions of 50% for the two populations ($f_a$=$f_b$=0.5) ii) on all the parameters, i.e. 3 Uncertain Parameters (3UP): the two viral decay rates and the initial proportions for each population.

Robust design optimization is performed following the FIM based HClnD-criterion (see Methods), following a combinatorial procedure over the 27132 possible designs of 7 sampling times. The initial time (t = 0 d) is included in each possible design and the 6

other samplings are optimized among the following vector of possible times: (1, 2, 3, 4 , 5, 6, 8, 10, 13, 16, 20, 24, 28, 32, 36, 40, 45, 50, 60) d.

We compare these methods and scenarios with two other design procedures: a non-optimized (equispaced) design and the local optimal design (i.e. optimal design if we know all the parameter values). Accounting for the different parameter possibilities (decay rates between 0.02 and 0.5 /d and initial proportions between 0.05 and 0.95) as in heatmaps (FIG 2 for non-optimized and local optimal design, Suppl. FIG 1 for robust designs), the coverage is defined as the proportion of parameter possibilities that are covered to obtain a certain precision of these parameter estimations or better, i.e. a RSE < 30% or 50% (FIG 4) in function of the design strategy. The allocation of sampling times for the robust designs was (0, 4, 5, 16, 20, 40, 60) days in case of 1 parameter uncertainty and (0, 4, 10, 20, 32 ,50, 60) days in the 3 parameter uncertainty scenario.

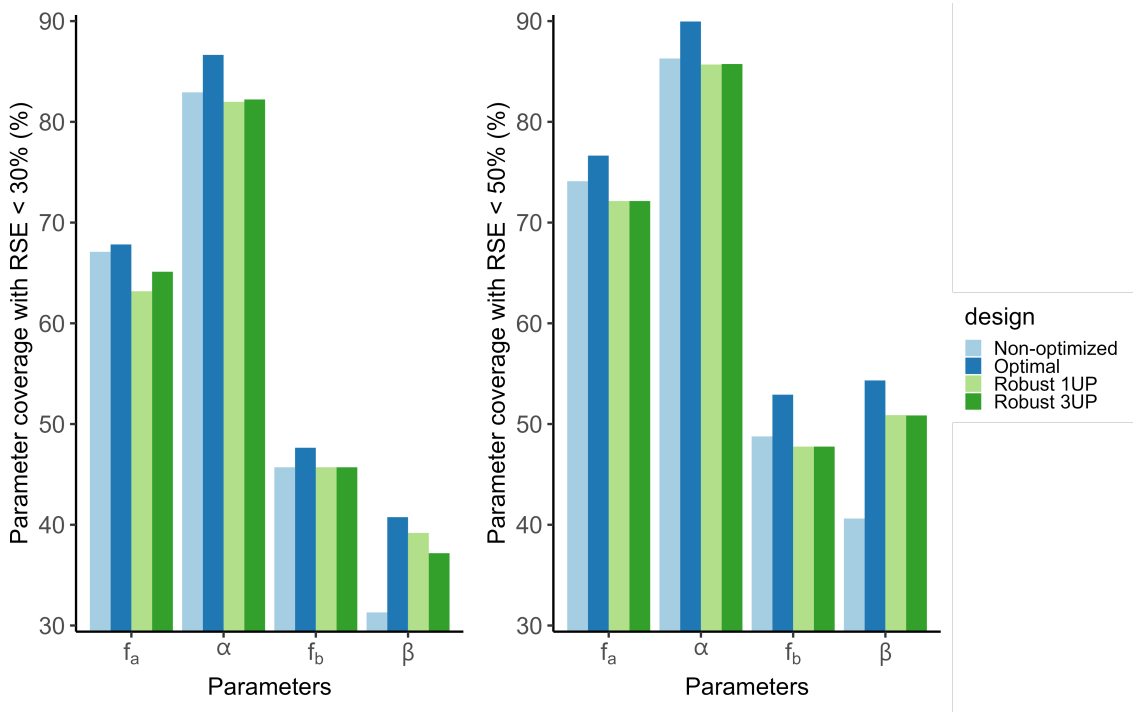



**FIG 4 Design comparison for expected coverage of parameter precision.** Coverage is the percentage of parameter settings ($\alpha$ from 0.02 to 0.48 /d (increment of 0.02), $\beta$ (>$\alpha$) from 0.04 to 0.5 /d, $f_a$ and $f_b$ among (0.05, 0.25, 0.5, 0.75 or 0.95, with $f_a$+$f_b$=1) that the different design strategies allow to reach an

expected RSE (Relative Standard Error) of 30% (left) or 50% (right) for the different parameters.

Parameters relative to the virus with faster elimination ($f_b$ and $\beta$) were the most difficult to estimate precisely. For instance, reaching a good RSE of 30% or acceptable RSE of 50% on the estimation of $\beta$ was covered in respectively 31% or 41% of cases if the design was not optimized. Due to early time points (5 in the first 20 days for Robust 1UP and 4 for Robust 3UP vs. 1 point every 10 days in the non-optimized design), the robust designs allowed to reach coverages of 39% (1UP) and 37% (3UP) or 51% (1UP and 3UP), respectively. Note that the performances of robust designs in case of parameter uncertainties are close to the coverage values of optimal designs (41% with RSE below 30% or 54% with RSE below 50%) if we would know the parameters at the design step, while a robust design is unique and optimal designs are calculated for each possible parameter vector. Coverages were however better and comparable regardless of the design for the estimation precision of parameters relative to the slower virus ($f_a$ and $\alpha$). The key point is that optimal design was able to significantly improve the fraction of contexts in which the fast decay rate $\beta$ was precisely estimated compared to non-optimized designs and even in the face of parameter uncertainty (see right-most columns, in FIG 4).

In this work, assuming uncertainty on three parameters results in similar performances as for one parameter (and considering the other parameters to be known). However, robust design methods can avoid a local optimal design that could be inadequate if the real values are very different compared to the prior values. Therefore, this method is preferable in a situation where we do have uncertainty in all parameters.

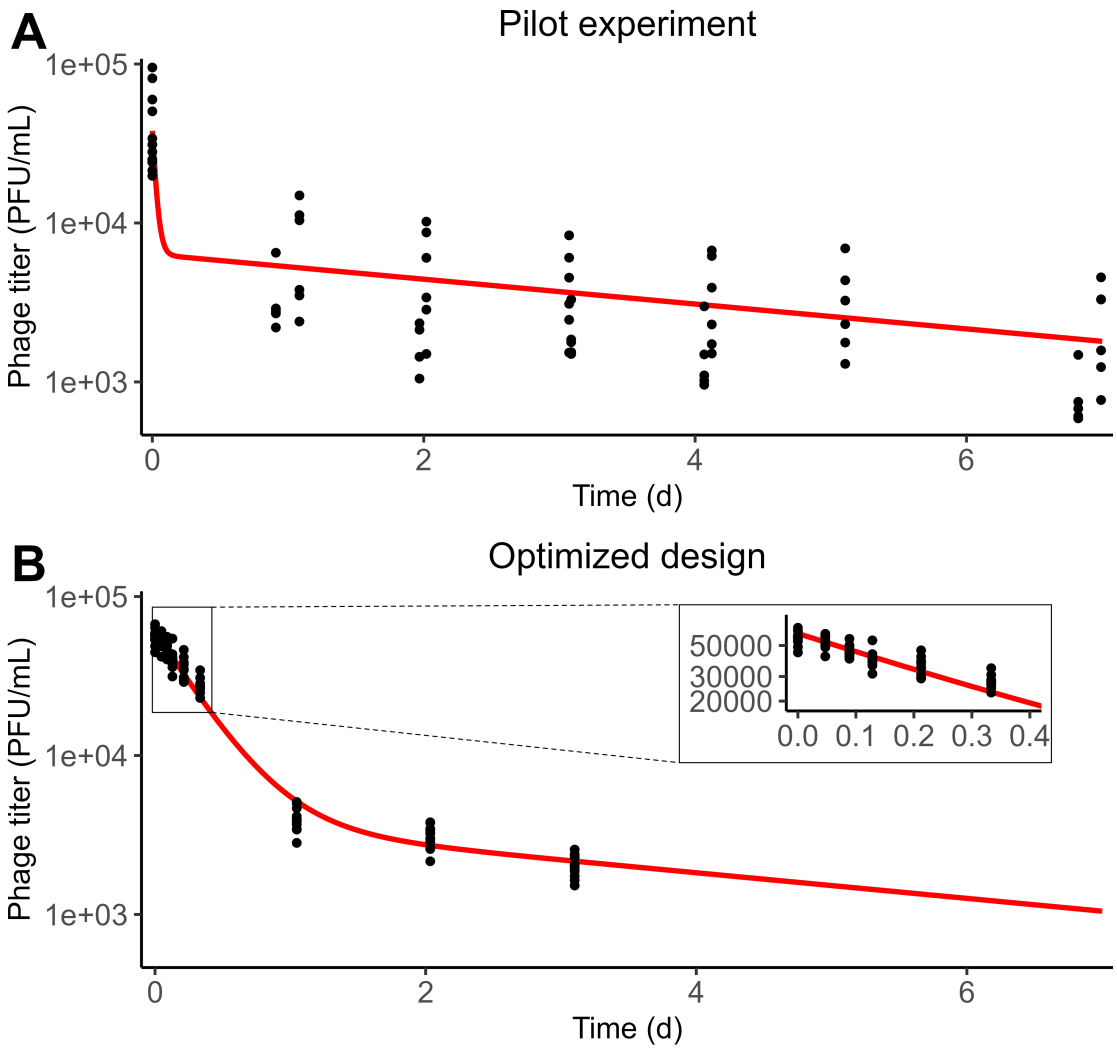
## Application: design optimization of phage ΦD9 decay

A first study of the phage ΦD9 decay (a pilot experiment) had been performed, using a non-optimized empirical design of 7 days with one measurement each day (every 24 h) excepted at day 6. The data showed evidence of a multiphasic decay (FIG 5A) but data fitting with the biexponential model (eq. 6, in that case, $V_0$ was estimated as well as $f_a$ and $f_b$ was fixed as $f_b$=$f_a$-1) showed that a precise estimation of the first (fastest) phase of this

decay ($\beta$) was not possible (TABLE 2) due to the rapid decay of the first subpopulation occuring during the first day without sufficient sampling.

Therefore, our design aim was to propose revised sampling times for a follow-up experiment allowing a precise estimation of $\beta$ and other parameters. An additional objective was to propose a design that could allow to identify whether there was triphasic decay, comprised of two highly unstable subpopulations that both decayed during the first day. Given the knowledge from the first experiment on the slower subpopulation and its initial density, we defined the following design constraints: a total of 9 sampling times including 0, 1, 2 and 3 days, and optimizing the 5 other sampling times located between 0 and 1 day.

Since biphasic and triphasic are nested models (see equation of triphasic model in Supplementary material), robust design optimization was performed for the most complex model, i.e. the triphasic decay model, assuming that a design enabling to estimate three phases of viral decay is also adequate to estimate two phases if only two are observable. From the inference of the pilot experiment, we assumed $\alpha$, $f_a$ and $V_0$ as well-constrained prior parameters (see values in TABLE 2). Using the HClnD-criterion based optimization, we accounted for uncertainty on the parameters relative to the two viral subpopulations with fastest rate. The robust design was composed of the following sampling times at 0, 0.04, 0.08, 0.13, 0.21, 0.33, 1, 2 and 3 days.




358

**FIG 5 ΦD9 decay data fitting from (A) the non-optimized pilot experiment and (B) the optimized design**. Red line is the viral density decay prediction from the biexponential model (eq. 2) and the parameters given in TABLE 2. The pilot experiment was composed of two batches with theoretical sampling times at 0, 1, 2, 4, 5, and 7 days for one batch and 0, 1, 2, 4, and 7 days for the other batch. Six replicates were made for each measurement of the pilot experiment. The optimized design experiment was composed of one batch with theoretical sampling times at 0, 0.04, 0.08, 0.13, 0.21, 0.33, 1, 2, and 3 days (i.e. 0, 1, 2, 3, 5, 8, 24, 48, and 72 hours). Twelve replicates were made for each measurement of the optimized experiment.

**TABLE 2 Parameter estimates from the different designs of ΦD9 decay.** The vector of sampling times is given in days (d), $\alpha$ is the decay rate of the slowest

virus subpopulation, $\beta$ is the decay rate of the fastest subpopulation, $V_0$ is the initial total virus density, $f_a$ and $f_b$ are the initial proportions (corresponding to $\alpha$ and $\beta$, respectively), $\sigma$ is the proportional error parameter, RSE is the relative standard error: a RSE greater than 100% is indicated by dark grey cell, and hyphen is indicated if the parameter was fixed (i.e. not estimated).

| | **Non-optimized (0, 1, 2, 3, 4, 5, 7) d** | | **Optimized (0, .04, .08, .13, .21, .33, 1, 2, 3) d** | |
|---|---|---|---|---|
| **Parameter (unit)** | **Estimate** | **RSE(%)** | **Estimate** | **RSE(%)** |
| $\alpha$ (/d) | .18 | 25 | .19 | 36 |
| $\beta$ (/d) | 39 | NaN | 3.2 | 6 |
| $V_0$ | $3.8\ 10^4$ | 20 | $6.1\ 10^4$ | 3 |
| $f_a$ | .17 | 28 | .06 | 17 |
| $f_b$ (=1-$f_a$) | .83 | - | .94 | - |
| $\sigma$ | .70 | 8 | .17 | 7 |

Data collected from this second experiment, with optimized sampling times, enables precise estimates of the parameters of the biphasic decay model (eq. 6), including the faster decay phase, despite a higher initial density (FIG 5B, TABLE 2). Indeed, the first decay rate $\beta$ (our main objective), wrongly estimated to a very high value (39 /d with RSE = NaN) after the non-optimized pilot experiment, is now identified and estimated at 3.2 /d, with a high precision (RSE = 6%). The estimation of the slower decay rate $\alpha$ is re-estimated to 0.19 /d with a moderate precision (RSE = 36%), which is consistent with the value obtained from the first experiment of 0.18 /d. We can also observe a lower noise than expected (the error parameter $\sigma$ was equal to 0.17 compared to 0.70 from the fitting of the first experiment).

A triphasic model with a faster initial decay did not improve the fit: AIC was lower for the biphasic model: 2045.7 vs. 2049.7 for the triphasic model, i.e. the biphasic model is more parsimonious. Even though our experiment included 7 samplings on the first day, we were unable to detect more than two subpopulations (see Supplementary material). It

is possible that multiple subpopulations exist with similar decay rates, however alternative approaches would be required to identify them.

## DISCUSSION

Here we revisited classic assumptions of exponential decay of virus particles to evaluate, optimize, and implement an alternative approach to measuring virus densities over time. In doing so, we show when and how an information-theoretic method can be used to leverage prior information and construct practical experimental designs that can accurately infer the proportion of viral subpopulations and their corresponding decay rates. We then successfully implemented an information-theoretic design to infer both a slow and fast decay rate in phage ΦD9, outperforming conventional equispaced design schemes.

Here, subpopulations of phages are initially defined using simple models with one decay rate corresponding to one subpopulation, whereas it has been shown that bacteria-phage co-evolution can have an impact on proper viral properties including decay rate [21, 4]. Also, even if a viral subpopulation is considered, the reality is more complex as there is likely a distribution of decay rates inside one population. Indeed, a good design choice helps to identify and characterize different subpopulations that can correspond to different mutations and possibly enhanced adsorption of a bacteriophage [21]. Optimal design can also be compatible with other statistical models such as mixed effect models [8]. Importantly, optimal design methods presented in this article can also be used in other fields such as studies on the biphasic variation of decay rate with temperature [14], the decay dynamics of HIV [26] or bacterial evolution and dormancy [5].

The implementation of an information theoretic design scheme provides a route to distinguish biphasic from monophasic decay. However, we recognize that there are limits to such inference given uncertainty in the underlying decay model. Alternative methods exist to perform design optimization in case of structural model uncertainty [17, 25]. In our real application, data from the two experiments were fitted separately, the first one being only used to design the second experiment. One potential alternative would have

been to pool the data from multiple experiments to improve parameter precision of estimation, as in model-based adaptive designs [22, 9]. Nonetheless, we leveraged the first experiment to establish a prior for accurately estimated parameters and not for those with significant uncertainty - modulated in part by the impact of noise on inference. We caution that noise has a linear impact on parameter imprecision such that measurement error propagates linearly to impact RSE - a significant issue in estimating rapid decay.

Moving forward, we anticipate the value of exploring alternative optimal experiment design methods as part of future experiments. These experimental designs could include ElnD-criterion based robust designs which accounts for all the expected distribution of uncertain parameters but are more time-consuming to compute relative to than HClnD-designs [12]. In such cases, the design choice can also be influenced by priorities, e.g., choosing to focus on estimating specific parameters of a biphasic decay, such as the fastest decay rate. For this purpose, the $DD_S$-criterion can be used at it allows an optimal balance between the parameter(s) of interest and other parameters of the models [2]. A cost function accounting for the number of samples can also be included during design optimization. Finally, alternative optimization algorithms may also be appropriate, notably when trying to infer parameters associated with decay in the context of more complex models [28, 15].

To conclude, this study propose methods to design viral decay experiments and shows the conceptual and practical value of design evaluation and optimization. By leveraging prior information and a pilot experiment, we were able to design an improved, follow-up experiment that robustly inferred the presence of a large majority of fast-decay viruses within the ΦD9 population. Beyond the practical value of improving specific parameter estimates, the use of robust design may also open the door to identifying the genetic and/or epigenetic basis for variation in viral life history traits.

## FUNDING

This work was supported by the following: Howard Hughes Medical Institute Emerging Pathogens Initiative grant 7012574 (J.R.M.), National Institutes of Health grants

T32GM007240 (J.M.B.) and 1R01AI46592-01 (J.S.W.), and the Chaire Blaise Pascal program of the Île-de-France region (J.S.W.).

# REFERENCES

## References

[1] Atkinson, A., Donev, A., and Tobias, R. *Optimum experimental designs, with SAS*. Oxford, New York: Oxford University Press, 2007.

[2] Atkinson, A. C., and Bogacka, B. Compound d-and ds-optimum designs for determining the order of a chemical reaction. *Technometrics 39*, 4 (1997), 347–356.

[3] Blazanin, M., Lam, W. T., Vasen, E., Chan, B. K., and Turner, P. E. Decay and damage of therapeutic phage omko1 by environmental stressors. *PLoS One 17*, 2 (2022), e0263887.

[4] Borin, J. M., Lee, J. J., Lucia-Sanz, A., Gerbino, K. R., Weitz, J. S., and Meyer, J. R. Rapid bacteria-phage coevolution drives the emergence of multiscale networks. *Science 382*, 6671 (2023), 674–678.

[5] Brouwer, A. F., Eisenberg, M. C., Remais, J. V., Collender, P. A., Meza, R., and Eisenberg, J. N. Modeling biphasic environmental decay of pathogens and implications for risk analysis. *Environmental science & technology 51*, 4 (2017), 2186–2196.

[6] De Paepe, M., and Taddei, F. Viruses' life history: towards a mechanistic basis of a trade-off between survival and reproduction among phages. *PLoS biology 4*, 7 (2006), e193.

[7] Dean, K., and Mitchell, J. Identifying water quality and environmental factors that influence indicator and pathogen decay in natural surface waters. *Water Research 211* (2022), 118051.

[8] Dumont, C., Lestini, G., Le Nagard, H., Mentré, F., Comets, E., Nguyen, T. T., et al. Pfim 4.0, an extended r program for design evaluation and optimization in nonlinear

mixed-effect models. *Computer methods and programs in biomedicine 156* (2018), 217–229.

[9] Fayette, L., Leroux, R., Mentré, F., and Seurat, J. Robust and adaptive two-stage designs in nonlinear mixed effect models. *The AAPS Journal 25*, 4 (2023), 71.

[10] Fedorov, V. V. *Theory of optimal experiments*. Elsevier, 2013.

[11] Fedorov, V. V., and Leonov, S. L. *Optimal design for nonlinear response models*. CRC Press, 2013.

[12] Foo, L. K., McGree, J., Eccleston, J., and Duffull, S. Comparison of robust criteria for d-optimal designs. *Journal of biopharmaceutical statistics 22*, 6 (2012), 1193–1205.

[13] Guedj, J., Thiébaut, R., and Commenges, D. Practical identifiability of hiv dynamics models. *Bulletin of mathematical biology 69* (2007), 2493–2513.

[14] Hiatt, C. Kinetics of the inactivation of viruses. *Bacteriological reviews 28*, 2 (1964), 150–163.

[15] Kennedy, J., and Eberhart, R. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks* (Perth, WA, Australia, 1995), vol. 4, IEEE, pp. 1942–1948.

[16] Kuhn, E., and Lavielle, M. Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis 49*, 4 (2005), 1020–1038.

[17] Loingeville, F., Nguyen, T. T., Riviere, M.-K., and Mentré, F. Robust designs in longitudinal studies accounting for parameter and model uncertainties–application to count data. *Journal of Biopharmaceutical Statistics 30*, 1 (2020), 31–45.

[18] Mentre, F., Mallet, A., and Baccar, D. Optimal design in random-effects regression models. *Biometrika 84*, 2 (1997), 429–442.

[19] Nelder, J. A., and Mead, R. A simplex method for function minimization. *The computer journal 7*, 4 (1965), 308–313.

[20] Nguyen, T. T., Bazzoli, C., and Mentré, F. Design evaluation and optimisation in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Statistics in Medicine 31*, 11-12 (2012), 1043–1058.

[21] Petrie, K. L., Palmer, N. D., Johnson, D. T., Medina, S. J., Yan, S. J., Li, V., Burmeister, A. R., and Meyer, J. R. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science 359*, 6383 (2018), 1542–1545.

[22] Pierrillas, P. B., Fouliard, S., Chenel, M., Hooker, A. C., Friberg, L. F., and Karlsson, M. O. Model-based adaptive optimal design (mbaod) improves combination dose finding designs: an example in oncology. *The AAPS journal 20* (2018), 1–11.

[23] Pronzato, L., and Pázman, A. *Design of experiments in nonlinear models*. New York: Springer Science Business Media, 2013.

[24] Richter, Ł., Księżarczyk, K., Paszkowska, K., Janczuk-Richter, M., Niedziółka-Jönsson, J., Gapiński, J., Łoś, M., Hołyst, R., and Paczesny, J. Adsorption of bacteriophages on polypropylene labware affects the reproducibility of phage research. *Scientific reports 11*, 1 (2021), 7387.

[25] Seurat, J., Nguyen, T. T., and Mentré, F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *Statistical methods in medical research 29*, 3 (2020), 934–952.

[26] White, J. A., Simonetti, F. R., Beg, S., McMyn, N. F., Dai, W., Bachmann, N., Lai, J., Ford, W. C., Bunch, C., Jones, J. L., et al. Complex decay dynamics of hiv virions, intact and defective proviruses, and 2ltr circles following initiation of antiretroviral therapy. *Proceedings of the National Academy of Sciences 119*, 6 (2022), e2120326119.

[27] Wynn, H. P. Results in the theory and construction of d-optimum experimental designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology 34*, 2 (1972), 133–147.

[28] Yu, Y. Monotonic convergence of a general algorithm for computing optimal designs. *The Annals of Statistics 38*, 3 (2010), 1593–1606.