COMPARING DATA COLLECTION STRATEGIES VIA INPUT UNCERTAINTY WHEN SIMULATING TESTING POLICIES USING VIRAL LOAD PROFILES

Drupad Parmar
Lucy E. Morgan
Andrew C. Titman

STOR-i Centre for Doctoral Training
Lancaster University
Lancaster, LA1 4YW, UK

Eva D. Regnier

Graduate School of Defense Management
Naval Postgraduate School
Monterey, CA 93943, USA

Susan M. Sanchez

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, USA

ABSTRACT

Temporal profiles of viral load have individual variability and are used to determine whether individuals are infected based on some limit of detection. Modelling and simulating viral load profiles allows for the performance of testing policies to be estimated, however viral load behaviour can be very uncertain. We describe an approach for studying the input uncertainty passed to simulated policy performance when viral load profiles are estimated from different data collection strategies. Our example shows that comparing the strategies solely based on input uncertainty is inappropriate due to the differences in confidence interval coverage caused by negatively biased simulation outputs.

1 INTRODUCTION

Individuals infected by a virus have a temporal profile of viral load starting from the moment they are infected. If and when they are tested their test sensitivity, the probability that an infected individual will correctly receive a positive test result, is a function of their viral load. There is individual variability in profiles of viral load, and therefore in test sensitivity too. Viral load can be a chief predictor for the risk of virus transmission (Quinn et al. 2000) and can also be strongly linked to mortality (Pujadas et al. 2020), therefore monitoring and understanding the behaviour of viral load is extremely important. We are interested in modelling the individual variability in temporal profiles of viral load and test sensitivity, and using this model to simulate and compare the performance of different testing policies, where a testing policy is a choice of how many times to test an individual and when. This might assist in suppressing the spread of a virus by for example indicating which testing policies are most effective in terms of successfully identifying the virus. Estimates of the shape of viral load profiles and the individual variability can be very uncertain since it requires observations of viral load across a group of individuals at different time points. In collecting such data there are choices to be made regarding the number of individuals to observe, the number of times to observe each individual and the time points at which the observations are taken. We shall refer to a particular choice of these as a data collection strategy. Modelling viral load profiles

directly from data will introduce a source of uncertainty into the simulation that will propagate through to the outputs. We wish to study how data collection strategies differ in terms of the amount of uncertainty they pass through the simulation to the estimated performance of a testing policy, as if a particular strategy offers a reduced level of uncertainty then this can be used for future data collection.

Here we outline an approach to compare the uncertainty passed to simulation outputs due to different data collection strategies. We use a nonlinear mixed effects model to allow for individual variability in viral load profiles and test sensitivity, and using this as our input model we are able to simulate the performance of different testing policies. We focus on how different data collection strategies used to estimate the input model parameters compare in terms of the uncertainty that propagates through the simulation model to the output, known as input uncertainty. This comparison allows us to consider whether a particular strategy should be adopted in order to obtain a reduced level of input uncertainty and thus gain greater insight from the simulated performance of testing policies.

The rest of the paper is organised as follows, we discuss background literature in Section 2 and outline the steps for our general approach in Section 3. An experiment using this approach is illustrated in Section 4. We discuss the results of our experiment and areas for future work in Section 5 before concluding in Section 6.

2 BACKGROUND

Lindstrom and Bates (1990) proposed a general nonlinear mixed effects model for data observed from a number of individuals repeatedly under different conditions, otherwise known as repeated measures data. They estimate the model parameters via a two-step iterative procedure that draws on methods developed for nonlinear fixed effects models and linear mixed effects models.

Chen et al. (2020) model respiratory viral load for patients infected with SARS-CoV-2 using a mechanistic model for respiratory virus kinetics based on a system of differential equations. Heterogeneity of viral load is evaluated by fitting Weibull distributions at each time point which sample data is available. Kucirka et al. (2020) model the time profile of test sensitivity using Bayesian hierarchical logistic regression. By using a nonlinear mixed effects model we are able to incorporate individual variability in viral load and test sensitivity under a single model.

Quilty et al. (2021) consider how quarantine and testing policies impact transmission of SARS-CoV-2 using an agent-based model to simulate the dynamics of viral load for exposed individuals. Larremore et al. (2021) use temporal profiles of viral load to simulate the effectiveness of certain testing policies and tests with different properties in controlling an epidemic. Rather than aiming to compare different testing policies directly, we are interested in how using different data collection strategies to model viral load profiles can result in varying levels of uncertainty in simulated responses.

Barton (2012) and Song et al. (2014) give an introduction to input uncertainty, a term which refers to the variance passed to a simulation output due to having estimated the input models via real-world data, and describe techniques to quantify it. For recent advancements see Lam and Qian (2017), Lam and Qian (2018). Efforts by Nelson et al. (2020) have also been made to reduce input uncertainty using frequentist modelling averaging. Morgan et al. (2019) recently considered detection and quantification of bias caused by input modelling.

3 APPROACH

In this section we outline the steps in our approach. We introduce our model for viral load and discuss how we simulate the performance of testing policies. We define input uncertainty, a method for quantifying it, and describe how we can compare the impact different data collection strategies have on it.

3.1 Modelling Viral Load

To represent the individual variability in temporal profiles of viral load we use a nonlinear mixed effects model, as it seems appropriate that all profiles follow the same functional form but with individual variability in some parameters. Let us consider a group of infected individuals who have unique temporal profiles of viral load and suppose we are able to observe, with noise, the viral loads of these individuals at different time points. An example of this can be found in Wölfel et al. (2020), where daily measurements of viral load are taken from different patients using RT-PCR. Let y_{ij} denote the viral load observed in individual i at time t_{ij}, where $i = 1, \ldots, N$, and $j = 1, \ldots, n_i$. Our nonlinear mixed effects model is as follows

$$y_{ij} = v(t_{ij}, \boldsymbol{\theta}_i) + g(t_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\phi}) \varepsilon_{ij},$$
$$\boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{\theta}_{\text{pop}}, \boldsymbol{\Omega}),$$
$$\varepsilon_{ij} \sim \mathcal{N}(0, 1),$$

where v is a nonlinear function for the temporal profile of viral load, $\boldsymbol{\theta}_i$ is a parameter vector of length m for the ith individual, g is a function for the residual error, $\boldsymbol{\phi}$ is a parameter vector of length q for the residual error function and ε_{ij} is an error term which follows a standard normal distribution. Individual parameter vectors are assumed to be normally distributed with mean $\boldsymbol{\theta}_{\text{pop}}$, a vector of length m representing the fixed population parameters and variance $\boldsymbol{\Omega}$, an $m \times m$ covariance matrix representing the random effects. The residual errors are assumed to be independent for different individuals and independent of each other for the same individual.

The parameters of the model $(\boldsymbol{\theta}_{\text{pop}}, \boldsymbol{\Omega}, \boldsymbol{\phi})$ can be estimated by maximising the log-likelihood function or the restricted log-likelihood function of the observed data. These expressions cannot be evaluated analytically as they have no closed form, however computational methods can be used to obtain parameter estimates $(\hat{\boldsymbol{\theta}}_{\text{pop}}, \hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\phi}})$.

3.2 Simulation

We can use the estimated fixed and random effects to simulate a large number of parameter vectors, where each parameter vector defines the temporal profile of viral load for an individual. This tells us about the behaviour of viral load profiles amongst a population. Moreover, the residual error function and its estimated parameters tell us about the distribution of error that would occur in observing the viral load of an individual at any time point. A test result for an individual at a specific time is based on whether their viral load measurement lies above or below some limit of detection. For each simulated individual we can calculate the probability of their observed viral load lying above or below the limit of detection at some given testing time by considering the distribution of error around the true viral load.

Suppose we use the estimated fixed and random effects to simulate K individuals. Each individual k is defined by a parameter vector $\tilde{\boldsymbol{\theta}}_k$, where

$$\tilde{\boldsymbol{\theta}}_k \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_{\text{pop}}, \hat{\boldsymbol{\Omega}}).$$

The true viral load for individual k at some time point t is given by $v(t, \tilde{\boldsymbol{\theta}}_k)$, however we observe the viral load with noise. This noise is distributed according to the residual error function multiplied by a standard normal random variable. The residual error function can be estimated by $g(t, \tilde{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}})$, therefore it follows that the distribution of the observed viral load for individual k at time point t is

$$\mathcal{N}(v(t, \tilde{\boldsymbol{\theta}}_k), g(t, \tilde{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}})^2).$$

For some limit of detection γ, we can use this distribution to measure the probability of an individual correctly or incorrectly returning either a positive or negative test result at any time point. We can use the simulated individuals to study the performance of different testing policies amongst a large population.

The performance of a policy, which is a choice of how many times to test each individual and when, could be measured in different ways, for example using the expected number of true positive tests or the average probability of a false negative test.

Let $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_{\text{pop}}, \hat{\boldsymbol{\Omega}}, \hat{\boldsymbol{\phi}}\}$ denote the estimated parameters of the nonlinear mixed effects model that will drive the simulation. For a testing policy p we denote the simulation output of replication j by

$$Y_j(\hat{\boldsymbol{\Theta}}, p) = \eta(\hat{\boldsymbol{\Theta}}, p) + e_j(\hat{\boldsymbol{\Theta}}, p),$$

where $\eta(\hat{\boldsymbol{\Theta}}, p)$ is the expected value of the simulation output for policy p given the parameters $\hat{\boldsymbol{\Theta}}$, and e is a random variable with mean 0 representing stochastic noise. Note that within a replication a choice for the number of individuals to simulate must be made. We consider this to be analogous to choosing the run length of replications and so it is not included in the notation.

We assume that there are unknown but true parameters $\boldsymbol{\Theta}^0 = \{\boldsymbol{\theta}_{\text{pop}}^0, \boldsymbol{\Omega}^0, \boldsymbol{\phi}^0\}$ from which the real-world data is observed and that the goal of the simulation experiment is to estimate $\eta(\boldsymbol{\Theta}^0, p)$, the expected value of the simulation output for policy p given the true parameters $\boldsymbol{\Theta}^0$. We estimate this value by taking the sample mean across n simulation replications using the fitted parameters, that is

$$\bar{Y}(\hat{\boldsymbol{\Theta}}, p) = \frac{1}{n} \sum_{j=1}^{n} Y_j(\hat{\boldsymbol{\Theta}}, p),$$

$$= \eta(\hat{\boldsymbol{\Theta}}, p) + \frac{1}{n} \sum_{j=1}^{n} e_j(\hat{\boldsymbol{\Theta}}, p).$$

The variance of this estimator breaks down into two distinct terms, stochastic estimation error and input uncertainty. Stochastic estimation error arises from the random variates generated in each replication and can be easily estimated via the sample variance.

3.3 Input Uncertainty

Input uncertainty refers to the variance in the expected simulation output that arises due to having estimated the input models. In our case input uncertainty simply reduces to parameter uncertainty and is given by

$$\sigma_I^2 = \text{Var}[\eta(\hat{\boldsymbol{\Theta}}, p)].$$

Input uncertainty is not straightforward to estimate since it requires knowledge of the sampling distribution of $\hat{\boldsymbol{\Theta}}$ and the functional form of $\eta(\cdot)$ both of which are usually unknown (Song et al. 2014).

A simple way to estimate input uncertainty is using bootstrapping, which involves sampling with replacement from the input model data, fitting input models according to this bootstrapped data and running replications using the bootstrap fitted input models. Repeating this bootstrapping procedure multiple times and treating the results as a random-effects model allows us to estimate input uncertainty (Nelson 2013). In the case of repeated measures data bootstrapping can be done by resampling with replacement from entire individuals, this is known as a case or paired bootstrap.

3.4 Comparing Data Collection Strategies

By choosing the functions v and g, and selecting some true parameters $(\boldsymbol{\theta}_{\text{pop}}^0, \boldsymbol{\Omega}^0, \boldsymbol{\phi}^0)$ we can simulate viral load observations from the nonlinear mixed effects model. This allows us to generate data sets of observations where we choose the number of individuals observed, as well as the number of observations per individual and the times at which they are observed. Using these data sets we can estimate input parameters, simulate the performance of different testing policies and quantify input uncertainty. This will allow us to investigate how different data collection strategies perform in terms of the uncertainty they propagate to the simulation outputs. If a particular approach clearly returns a reduced level of input uncertainty compared to any alternatives then using this data collection strategy will provide more insightful simulation results and allow for better comparisons to be made between different testing policies.

4 EXPERIMENT

In this section we describe an experiment using the approach outlined in Section 3. We select functions and parameters for the nonlinear mixed effects model and compare three different data collection strategies across two sets of testing policies.

4.1 Input Model

To provide an example of functions and parameters we use the recent literature on SARS-CoV-2. Note we are not aiming to accurately model viral load profiles that occurred during the COVID-19 pandemic, rather just choose a nonlinear mixed effects model that is somewhat representative of a real-world problem. To model the temporal profile of viral load we choose a shifted scaled lognormal function as this has a single peak and an unbounded right tail, with a shape similar to the right-skewed plots of viral load in Wölfel et al. (2020). Individual i is represented by parameter vector $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \alpha_i)$, where μ_i and σ_i are parameters to the lognormal distribution and α_i is the scale parameter. The viral load of individual i at time t is given by

$$v(t, \boldsymbol{\theta}_i) = \begin{cases} \frac{\alpha_i f(t-s, \mu_i, \sigma_i)}{f(\exp(\mu_i - \sigma_i^2), \mu_i, \sigma_i)}, & t \geq s \\ 0, & t < s \end{cases}$$

where $f(t, \mu_i, \sigma_i)$ is the probability density function of a lognormal distribution with parameters μ_i and σ_i, $s > 0$ represents the shift, and t is interpreted as days since infection. The mode of $f(t, \mu_i, \sigma_i)$ is given by $t = \exp(\mu_i - \sigma_i^2)$, therefore α_i represents the peak of the viral load profile for individual i. Note that v measures viral load in units of $\log_{10}$.

We find fixed and random effects, and a shift value such that the viral load profiles generated by the nonlinear mixed effects model match properties found in the literature, such as the peak viral load behaviour in Chen et al. (2020) and the time between infection and the peak viral load in Backer et al. (2020). The fixed and random effects are given (to three decimal places) by the following

$$\boldsymbol{\theta}^0_{\text{pop}} = (3.453, 1.415, 7.023), \qquad \boldsymbol{\Omega}^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.030 & 0.046 \\ 0 & 0.046 & 2.639 \end{bmatrix},$$

with shift $s = 1.5$.

We consider a multiplicative residual error function with parameter vector $\boldsymbol{\phi} = (\beta, \rho)$, where the residual error function for individual i at time t is given by

$$g(t, \boldsymbol{\theta}_i, \boldsymbol{\phi}) = \beta v(t, \boldsymbol{\theta}_i)^\rho.$$

We set the true residual error parameters to be $\boldsymbol{\phi}^0 = (0.25, 1)$, as this means the test sensitivity over time exhibits similar behaviour to that found in Kucirka et al. (2020).

4.2 Simulation

In our experiment we measure policy performance using the average probability of an individual returning at least one positive test. An individual will return a negative test result if their observed viral load is below some limit of detection γ. The probability of a simulated individual k with parameters $\tilde{\boldsymbol{\theta}}_k$ returning a negative test result at time t, where $t > s$, is given by

$$\Phi\left(\frac{\gamma - v(t-s, \tilde{\boldsymbol{\theta}}_k)}{g(t, \tilde{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}})}\right),$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Reported limits of detection of the PCR tests for SARS-CoV-2 have a wide range, we will use $\gamma = 4$.

A policy that involves testing every individual at d time points given by $t_1, \ldots, t_d$, is denoted by $p_{t_1, \ldots, t_d}$. For example p_3 denotes a policy where every individual is tested on day 3, whilst $p_{3,5}$ denotes a policy where every individual is tested on both day 3 and day 5. The probability that an individual returns at least one positive test for a particular policy is equal to 1 minus the probability that the individual returns a negative test result at all time points in the policy. For individual k, the probability of returning at least one positive test under the policy $p_{t_1, \ldots, t_d}$ is given by

$$1 - \prod_{i=1}^{d} \Phi \left(\frac{\gamma - v(t_i - s, \tilde{\boldsymbol{\theta}}_k)}{g(t_i, \tilde{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}})} \right).$$

We can consider how the policy $p_{t_1, \ldots, t_d}$ performs across all simulated individuals by computing the mean probability of an individual returning at least one positive test under the policy, thus our simulation output is

$$\frac{1}{K} \sum_{k=1}^{K} \left[1 - \prod_{i=1}^{d} \Phi \left(\frac{\gamma - v(t_i - s, \tilde{\boldsymbol{\theta}}_k)}{g(t_i, \tilde{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}})} \right) \right].$$

This performance measure might be used to help successfully identify infected individuals and implement effective isolation measures.

We simulate the performance of 22 policies which are split into two sets. The first set consists of policies $p_{t_1}, p_{t_1, t_2}, p_{t_1, t_2, t_3}$ where $t_2 = t_1 + 2$ and $t_3 = t_1 + 4$ for $t_1 = 3, 4, 5, 6$. These policies might be simulated to study how increased testing affects policy performance, we will refer to them as the multiple testing policies. The second set consists of policies p_3 and p_{3,t_2}, for $t_2 = 4, \ldots, 14$. These policies might be simulated to help decide when to conduct a follow up test on individual, we will refer to these as the follow up testing policies.

4.3 Data Collection Strategies

We are interested in comparing data collection strategies where the total number of viral load observations are the same but the observations are split differently in terms of the number of individuals observed and the number of observations per individual. We consider three strategies all of which use 350 observations of viral load. The first takes 10 observations from 35 individuals, the second takes 7 observations from 50 individuals, and the third takes 5 observations from 70 individuals. These will be referred to as the 35x10, 50x7, and 70x5 strategies respectively. For each strategy individuals are observed at equally spaced time points with their first observation taken at a randomised time after s such that their final observation is taken before $21 + s$.

4.4 Results

We now have our input model functions and parameters, simulation output, testing policies, and different data collection strategies to compare. For each strategy we generate a data set of viral load observations as appropriate from the true input model and estimate the input model parameters using the R package developed by Pinheiro et al. (2020) which implements the computational methods described in Lindstrom and Bates (1990). These input parameters drive the simulation model in the nominal experiment, where each policy performance is estimated across $n = 100$ replications, each of which simulates $K = 10000$ individuals. We then approximate input uncertainty and stochastic estimation error for each policy performance via a diagnostic experiment where the data is bootstrapped $B = 250$ times, and where for each bootstrap the same number of replications and individuals are used as in the nominal experiment. This gives us an estimate of policy performance, input uncertainty and stochastic estimation error for each of the 22 policies.

Measures of input uncertainty and stochastic estimation error are typically used to construct confidence intervals for simulation outputs. We use the asymptotic normality theory method described in Cheng and Holland (2004), however we make a slight adjustment. Since the simulation output is a probability we

construct 95% confidence intervals using normality theory on the logarithm of one minus the bootstrapped outputs, and then transform the intervals back to the original scale. The confidence intervals for each policy performance produced by the three data collection strategies are shown in Figure 1. Note that the stochastic estimation error is very small so the uncertainty in policy performance is primarily due to input uncertainty.
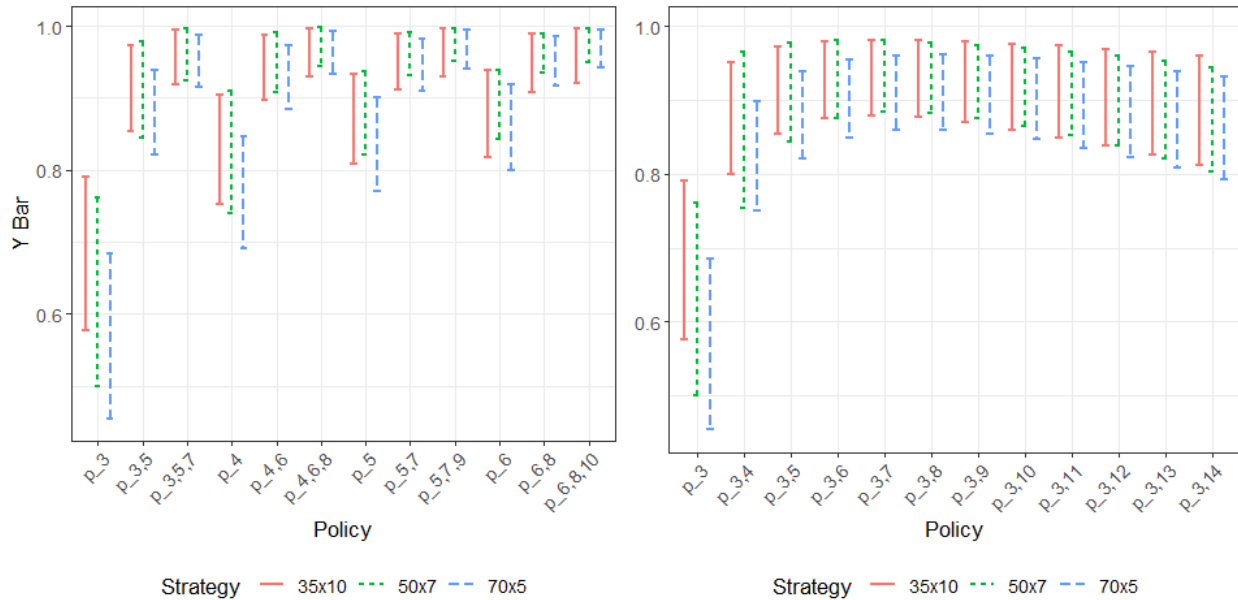

Figure 1: Confidence intervals produced by each data collection strategy in a single macro replication.

We can see that the widths of the confidence intervals and their estimates of the performance measure differ for each strategy. In this case the 35x10 strategy has the smallest input uncertainty for policies p_3, $p_{3,5}$, p_4 and $p_{3,4}$, whilst the 50x7 strategy has the smallest input uncertainty for all other policies. Note that policies that test more frequently have a reduced level of input uncertainty and this translates to narrower confidence intervals. To understand the behaviour of each strategy more generally we conduct 100 macro replications, that is we repeat this experiment 100 times. First we compare the average input uncertainty estimates for the two sets of policies across the three data collection strategies.

From Table 1 we see that all the multiple testing policies have the smallest average input uncertainty when using the 70x5 strategy and the largest average input uncertainty when using the 35x10 strategy. The same holds for all the follow up testing policies, apart from policy $p_{3,4}$ where the 50x7 strategy has the smallest average input uncertainty. Although not shown here, the average stochastic estimation error is of order $\times 10^{-7}$ for p_3 and order $\times 10^{-8}$ for all other policies across each of the data collection strategies, and is therefore relatively small compared to input uncertainty. Since input uncertainty and stochastic estimation error combine to give the total variance of a simulation output, these results suggest that the simulation outputs from using the 70x5 data collection strategy will have smallest variance and will hence offer the greatest level of insight.

As we have seen in Figure 1, the confidence intervals produced by each data collection strategy can cover quite different output values. Using the true input model parameters to run simulation replications we are able to estimate the true performance of each policy, and check whether these lie in the confidence intervals. Table 2 shows how many of the 100 macro replications produced confidence intervals that contained the true performance measure for each of the three data collection strategies.

For the multiple testing policies we see that generally the 50x7 strategy produces the most confidence intervals that contain the true performance measure. The 70x5 strategy produces many confidence intervals

Table 1: Average input uncertainty ($\times 10^{-4}$) for each data collection strategy across 100 macro replications.

(a) Multiple testing policies			
Policy	35x10	50x7	70x5
p_3	39.7	33.9	32.4
$p_{3,5}$	12.6	10.8	10.7
$p_{3,5,7}$	4.79	3.96	3.74
p_4	21.7	18.6	18.4
$p_{4,6}$	7.35	6.21	6.00
$p_{4,6,8}$	3.60	2.85	2.61
p_5	13.6	11.7	11.6
$p_{5,7}$	5.30	4.28	3.97
$p_{5,7,9}$	3.23	2.42	2.16
p_6	10.8	9.19	9.01
$p_{6,8}$	4.85	3.72	3.33
$p_{6,8,10}$	3.42	2.46	2.15

(b) Follow up testing policies			
Policy	35x10	50x7	70x5
p_3	39.7	33.9	32.4
$p_{3,4}$	21.6	18.6	18.9
$p_{3,5}$	12.6	10.8	10.7
$p_{3,6}$	9.03	7.68	7.43
$p_{3,7}$	7.79	6.50	6.17
$p_{3,8}$	7.55	6.20	5.82
$p_{3,9}$	7.82	6.36	5.93
$p_{3,10}$	8.40	6.79	6.30
$p_{3,11}$	9.19	7.40	6.85
$p_{3,12}$	10.1	8.16	7.54
$p_{3,13}$	11.2	9.03	8.35
$p_{3,14}$	12.3	10.0	9.27

Table 2: Number of macro replications out of 100 that produced confidence intervals containing the true performance measure for each data collection strategy.

(a) Multiple testing policies			
Policy	35x10	50x7	70x5
p_3	87	89	65
$p_{3,5}$	87	94	84
$p_{3,5,7}$	93	94	89
p_4	85	92	80
$p_{4,6}$	91	94	86
$p_{4,6,8}$	94	95	89
p_5	84	95	81
$p_{5,7}$	94	95	86
$p_{5,7,9}$	94	96	91
p_6	91	95	82
$p_{6,8}$	94	94	89
$p_{6,8,10}$	96	94	91

(b) Follow up testing policies			
Policy	35x10	50x7	70x5
p_3	87	89	65
$p_{3,4}$	90	94	82
$p_{3,5}$	87	94	84
$p_{3,6}$	87	94	82
$p_{3,7}$	90	93	84
$p_{3,8}$	91	92	84
$p_{3,9}$	93	91	85
$p_{3,10}$	93	92	87
$p_{3,11}$	93	89	86
$p_{3,12}$	92	86	86
$p_{3,13}$	92	87	86
$p_{3,14}$	92	87	85

that do not contain the true measure, particularly for the single day testing policies. For the follow up testing policies we again see that the 70x5 strategy produces the least amount of confidence intervals containing the true performance measure. For policies p_3 to $p_{3,8}$ the 50x7 strategy has the highest coverage whilst for policies $p_{3,9}$ to $p_{3,14}$ the 35x10 strategy has the highest coverage. These results suggest that the 50x7 approach performs best in terms of capturing the true performance measure. The results of Table 1 and Table 2 combined suggest that the 70x5 approach is reducing the variance the most but around the incorrect point, possibly indicating some input modelling bias. Note that across all policies the 35x10 strategy has the largest average interval width. The 70x5 strategy has the smallest average interval width for all policies apart from $p_{3,4}$ and $p_{3,5}$ where the 50x7 strategy has the smallest.

We investigate whether the true performance measure lies above or below the confidence intervals and find that there is a very strong tendency for the performance measure to lie above the confidence intervals

across all data collection strategies, particularly for the 50x7 and 70x5 strategy. This suggests that the policy performances estimated by the nominal experiments are underestimating the true performance measures, which we confirm by finding a negative bias for each simulated policy performance across every data collection strategy. Following this we calculate the bias in the input model parameter estimates. Figure 2 shows the relative bias induced by each data collection strategy across the 100 macro replications for each of the 8 input parameters. By considering relative bias we are able to make comparisons across parameters.
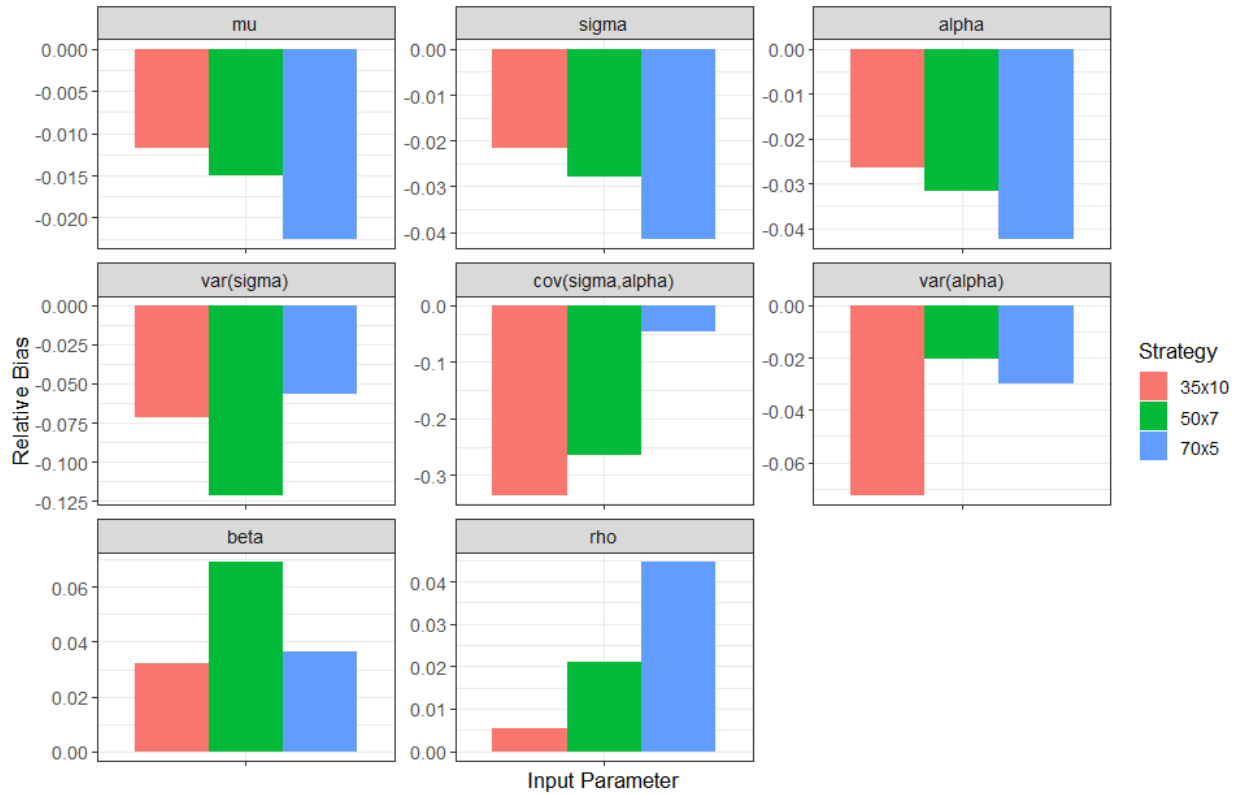

Figure 2: Relative input parameter bias across 100 macro replications for each data collection strategy.

From the top row we see that the relative bias in the population parameters, or the fixed effects, is smallest when using the 35x10 approach and is largest when using the 70x5 approach. Notably the relative bias is negative for each population parameter and strategy, perhaps indicating a tendency for each of these parameters to be underestimated regardless of how the data is collected. We see that parameters σ and α have a larger relative bias than μ for each strategy. Conversely the relative bias in estimating the terms of the variance matrix, or the random effects, generally seems to be largest when using the 35x10 strategy and smallest when using the 70x5 strategy. Again the relative bias is negative for each parameter and strategy, and is particularly large for the covariance of σ and α under the 35x10 and 50x7 strategy. The two terms in the residual error function have the smallest relative bias when using the 35x10 strategy.

Interpreting the results in Figure 2 is not straightforward as the relative bias in the estimated parameters occur simultaneously. For example, if we just consider a negatively biased estimate of α then this will clearly result in underestimated simulation outputs since this represents the mean peak height of the viral load profiles. However a biased estimate of α occurs at the same time as biased estimates of all other input parameters, and we do not know how these interact in terms of affecting the simulation output.

5 DISCUSSION

Initially we compared the input uncertainty induced by different data collection strategies and the results in Table 1 showed that the 70x5 data collection strategy generally had the smallest average input uncertainty across all policies. However the results in Table 2 showed that this strategy performs poorly in terms of producing confidence intervals for the simulation outputs that contain the true performance measure, whilst the 50x7 strategy showed the best coverage. We found that in many cases the true performance measures were greater than the upper bounds of the confidence intervals and that the true policy performance was being underestimated by the simulation outputs. Figure 2 showed that each data collection strategy provided negatively biased estimates of every fixed and random effect of the nonlinear mixed effects model.

Based on the results in this experiment it appears that the 50x7 strategy is most suitable, not because it returns a reduced level of input uncertainty across policies compared to the alternative strategies, but because it produces the best coverage of the true performance measure. Clearly it is inappropriate to solely compare the data collection strategies based on input uncertainty and that the confidence interval coverage should be considered, along with the bias in the simulation outputs. Should the data collection strategies give similar coverage and bias across all policies then comparing input uncertainty may help to differentiate between them.

With the example discussed we have seen negatively biased estimates of the fixed and random effects in our input model. Whether this is due to the amount of data being used to fit the input parameters or due to an inherent trait of the model fitting methodology remains to be seen and requires further investigation. Regardless of this, giving more consideration to how the input parameter estimates produced by each data collection strategy compare to the true input parameters may help us to understand the differences between strategies in terms of replicating the true input model behaviour. In addition, an input parameter sensitivity analysis would indicate which of the input model parameters is most influential in estimating each policy performance, as well as revealing how the simulation outputs are affected by interactions between the parameters.

More generally there are alternative methods for bootstrapping that could be used, such as the parametric bootstrap as well as bootstrap methods specific to nonlinear mixed effects models. Comparing these methods is difficult however as the true value of input uncertainty is unknown. There are also alternative techniques for quantifying input uncertainty that could be implemented which may offer a computational advantage over a bootstrapping procedure. We have seen issues with the simulation outputs behaving differently to the true performance measures, demonstrated by the poor confidence interval coverage shown by some of the data collection strategies. It therefore might be suitable to compare the strategies via the mean squared error due to input modelling, which incorporates input uncertainty and squared input model bias.

6 CONCLUSION

In this paper we outline an approach for comparing different data collection strategies via input uncertainty when modelling temporal profiles of viral load and simulating the performance of different testing policies. In particular, we use a nonlinear mixed effects model to represent individual variability in temporal profiles of viral load and use the residual error function to model measurement error. By selecting the functions and parameters of the model we can generate different sets of viral load observations that represent different strategies for collecting data. For each strategy we use the observations to estimate the model parameters which are used to simulate the performance of different testing policies. The different data collection strategies are compared by the impact of input uncertainty in the simulation outputs.

We demonstrate this approach using a nonlinear mixed effects model which aims to replicate the behaviour of viral load profiles seen during the COVID-19 pandemic. We simulate the probability of an individual returning at least one positive test across two sets of testing policies and compare data collection strategies which use the same number of total observations but split differently between the number of individuals and the number of observations per individual. In this example we find that comparing the

strategies solely based on input uncertainty is not sensible as the strategies perform differently in terms of producing confidence intervals for the simulation output that contain the true performance measure. We find a tendency for the true performance measures to be underestimated by the simulation and although this can be linked to negatively biased parameter estimates it requires further investigation through some sensitivity analysis. Future research includes considering alternative techniques for quantifying input uncertainty and incorporating confidence interval coverage and simulation output bias in the approach.

ACKNOWLEDGMENTS

REFERENCES

Backer, J. A., D. Klinkenberg, and J. Wallinga. 2020. "Incubation Period of 2019 Novel Coronavirus (2019-nCoV) Infections Among Travellers from Wuhan, China, 20–28 January 2020". *Eurosurveillance* 25(5):2000062.

Barton, R. R. 2012. "Tutorial: Input Uncertainty in Outout Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 1–12. Piscataway, New Jersey: IEEE.

Chen, P. Z., N. Bobrovitz, Z. Premji, M. Koopmans, D. N. Fisman, and F. X. Gu. 2020. "Heterogeneity in Transmissibility and Shedding SARS-CoV-2 via Droplets and Aerosols". https://www.medrxiv.org/content/early/2020/12/02/2020.10.13.20212233.full.pdf, accessed 26th April 2021.

Cheng, R. C., and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 14(4):344–362.

Kucirka, L. M., S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler. 2020. "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure". *Annals of Internal Medicine* 173(4):262–267.

Lam, H., and H. Qian. 2017. "Optimization-based Quantification of Simulation Input Uncertainty via Empirical Likelihood". https://arxiv.org/pdf/1707.05917.pdf, accessed 26th April 2021.

Lam, H., and H. Qian. 2018. "Subsampling to Enhance Efficiency in Input Uncertainty Quantification". https://arxiv.org/pdf/1811.04500.pdf, accessed 26th April 2021.

Larremore, D. B., B. Wilder, E. Lester, S. Shehata, J. M. Burke, J. A. Hay, M. Tambe, M. J. Mina, and R. Parker. 2021. "Test Sensitivity is Secondary to Frequency and Turnaround Time for COVID-19 Screening". *Science advances* 7(1):eabd5393.

Lindstrom, M. J., and D. M. Bates. 1990. "Nonlinear Mixed Effects Models for Repeated Measures Data". *Biometrics* 46:673–687.

Morgan, L. E., B. L. Nelson, A. C. Titman, and D. J. Worthington. 2019. "Detecting Bias due to Input Modelling in Computer Simulation". *European Journal of Operational Research* 279(3):869–881.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer.

Nelson, B. L., A. T. Wan, G. Zou, X. Zhang, and X. Jiang. 2020. "Reducing Simulation Input-Model Risk via Input Model Averaging". *INFORMS Journal on Computing*.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2020. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-144, https://CRAN.R-project.org/package=nlme.

Pujadas, E., F. Chaudhry, R. McBride, F. Richter, S. Zhao, A. Wajnberg, G. Nadkarni, B. S. Glicksberg, J. Houldsworth, and C. Cordon-Cardo. 2020. "SARS-CoV-2 Viral Load Predicts COVID-19 Mortality". *The Lancet Respiratory Medicine* 8(9):e70.

Quilty, B. J., S. Clifford, J. Hellewell, T. W. Russell, A. J. Kucharski, S. Flasche, W. J. Edmunds, K. E. Atkins, A. M. Foss, N. R. Waterlow et al. 2021. "Quarantine and Testing Strategies in Contact Tracing for SARS-CoV-2: A Modelling Study". *The Lancet Public Health* 6(3):e175–e183.

Quinn, T. C., M. J. Wawer, N. Sewankambo, D. Serwadda, C. Li, F. Wabwire-Mangen, M. O. Meehan, T. Lutalo, and R. H. Gray. 2000. "Viral Load and Heterosexual Transmission of Human Immunodeficiency Virus Type 1". *New England journal of medicine* 342(13):921–929.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 162–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wölfel, R., V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe et al. 2020. "Virological Assessment of Hospitalized Patients with COVID-2019". *Nature* 581(7809):465–469.

AUTHOR BIOGRAPHIES

DRUPAD PARMAR is a Ph.D. student at the Statistics and Operational Research Centre for Doctoral Training in Partnership with Industry at Lancaster University. His research is focused on input uncertainty quantification in stochastic simulation models. His email address is d.parmar1@lancaster.ac.uk. His website is https://www.lancaster.ac.uk/∼parmard1/.

LUCY E. MORGAN is a Lecturer in Simulation and Stochastic Modelling in the Department of Management Science at Lancaster University. Her research interests are input uncertainty in simulation models and arrival process modelling. Her e-mail address is l.e.morgan@lancaster.ac.uk.

EVA D. REGNIER is a Professor in the Graduate School of Defense Management at the Naval Postgraduate School. Her work seeks to better integrate information into operational decision-making, primarily by tailoring informational products for dynamic decision making. She holds Ph.D. and M.S. degrees from the Georgia Institute of Technology, and a B.S. from Massachusetts Institute of Technology. Her email address is eregnier@nps.edu.

SUSAN M. SANCHEZ is a Distinguished Professor of Operations Research at the Naval Postgraduate School, and Co-Director of the Simulation Experiments & Efficient Design (SEED) Center for Data Farming. She also holds a joint appointment in the Graduate School of Defense Management. She has a B.S. in Industrial & Operations Engineering from the University of Michigan, and a Ph.D. in Operations Research from Cornell. She has been an active member of the simulation community for many years, and has been recognized as a Titan of Simulation and an INFORMS Fellow. Her email address is ssanchez@nps.edu. Her web page is http://faculty.nps.edu/smsanche/.

ANDREW C. TITMAN received his Ph.D. from the University of Cambridge and currently is a Senior Lecturer in Statistics in the Department of Mathematics and Statistics at Lancaster University. His research interests include survival and event history analysis and latent variable modelling, with applications in biostatistics and health economics. His e-mail address is a.titman@lancaster.ac.uk.