AMSI VACATION RESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Statistical modelling of malaria parasite clearance rates

Meg Tully

Supervised by A/Prof Jennifer Flegg and Dr Sophie Zaloumis

The University of Melbourne

Vacation Research Scholarships are funded jointly by the Department of Education, Skills and Employment and the Australian Mathematical Sciences Institute.

Abstract

Background: Malaria is a mosquito-borne disease which caused over 400,000 deaths in 2019. The development of drug-resistant malaria strains has previously undermined front-line treatments, therefore the ability to systematically measure treatment effect is an important tool in monitoring the threat of resistance. A number of existing methods have been developed and three are examined in this project.

Methods: This project used Stan in an R interface to sample parameter values from an adapted version of the Bayesian hierarchical model presented in Fogarty et al 2015. This method represented patient-level parameters as drawn from population-level distributions with their own 'hyper parameters'. It also incorporated the 'lag' and 'tail' phases, where a patient's parasite density can remain constant for a period of time at the beginning and/or end of their profile.

Results: Sampling 4000 iterations (with 500 discarded as burn-in) from 4 chains took approximately 50 minutes. Diagnostic tools (such as Rhat values, Effective Ssample Size and trace plots) did not show any evidence of poor convergence or highly auto-correlated draws within each chain. Plots of 95% credible intervals illustrated the data seem to be well captured by the posterior samples. Estimates of covariate effects calculated from the posterior distributions were not in line with previous methods, indicating the need for further development of the model.

Conclusion: The methodology developed here is significantly faster, easier to implement and more flexible than existing methods. Unfortunately it is not yet accurate enough to be used to produce reliable estimates of covariate effects. Stan software is useful for increasing the accessibility of Bayesian sampling methods.

Introduction

Malaria was rated the 6th leading cause of death in low-income countries in 2019 and continues to disproportionately impact children and infants. Although treatments exist and are considered extremely effective, the development of resistant malaria strains threatens to significantly undermine efforts to achieve elimination (Menard & Dondorp 2017, Ouji et al 2018). In order to track and understand resistance, there needs to be systematic practices for quantitatively evaluating the effect of a treatment - in other words, a way to measure how well a drug (or combination therapy) neutralises malaria parasites.

A number of existing methods are currently used to evaluate treatment effect, this paper will briefly summarise these options in the context of a data set of 110 patients from Cambodia. A Bayesian modelling framework (created by Fogarty et al 2015) which balances group and individual level effects is then explored in more detail. Samples of parameter values from a version of this hierarchical model are simulated in the program Stan (Stan Development Team 2020) using an R interface RStan (R Core Team 2020).

Finally this report reviews the sampled parameter values and accompanying diagnostics, comparing the output to that of previous sampling methods. The comparison found this adapted Stan model to be more computationally efficient and much more flexible than methods available previously, however it is not yet able to sufficiently reproduce previous results for covariate effects.

Statement of Authorship

This report is a summary of work influenced and shaped by a large existing body of research on the modelling of malaria parasite clearance rates. This model presented here is an adaptation of that presented by Fogarty et al (2015), which is itself an expansion of the methods created for the Worldwide Antimalarial Resistance Network (Flegg et al 2011). The coding, sampling and report writing was done by Meg Tully. This work was completed under the guidance and supervision of A/Prof Jennifer Flegg and Dr Sophie Zaloumis.

BACKGROUND:

Malaria as a Global Health Challenge

Approximately 229 million malaria infections occurred in 2019, leading to 409,000 deaths, mostly among children under 5 (WHO 2020). Despite significant progress, the rate of mortality reduction has slowed since 2016 and more concerningly, current advancements are seriously threatened by emerging antimalarial resistance. Currently, the most effective treatment for malaria is artemisinin-based combination therapies (ACT). However, in Southeast-Asia, *Plasmodium falciparum* (the species responsible for most malaria deaths) is developing resistance to ACT treatments (Imwong et al 2017). In the past, effective antimalarial drugs have been repeatedly undermined by wide-spread resistance, which has tended to develop in Southeast Asia and spread into Africa.

The Dataset

The data analysed in this report is from the Pursat province of Cambodia (Amaratunga et al 2012). Artemisinin-resistant malaria is believed to have first emerged in Cambodia (Dondorp et al 2009, Hamilton et al 2019, Imwong et al 2017), therefore tacking treatment effect in this region is of particular interest. The data is from 110 patients infected with *Plasmodium falciparum* over 2009-2010. To be included a patient had to be over 10 years old, have an uncomplicated infection, an initial density over $10,000$ asexual parasites per μL, and a negative pregnancy test for women 15-45 years old. It's notable that 76.4% of presenting patients diagnosed with *P. falciparum* malaria during the study period were excluded due to these or other criteria. Each patient had parasite density/μL recorded every 6 hours until the densities became too low to be measured, any concentration below **15/mL** was recorded as 0. Therefore the number of measurements recorded for each patient vary significantly (min = 4, median = 13, max = 21). Mean parasite start count is 101,812 per μL (min = 11,882, max = 546,461).

ANALYSIS OPTIONS:

The goal of analysing parasite clearance data is twofold: firstly we aim to evaluate how effective a treatment is - and secondly examine the relationship between treatment impact and variables of interest. These could include geographical location, parasite genetic markers, dosage and treatment regime, patient demographics and date of treatment. By evaluating the year as a variable it can be possible to quantitatively define emerging resistance over time.

So what approaches can we use to achieve these goals?

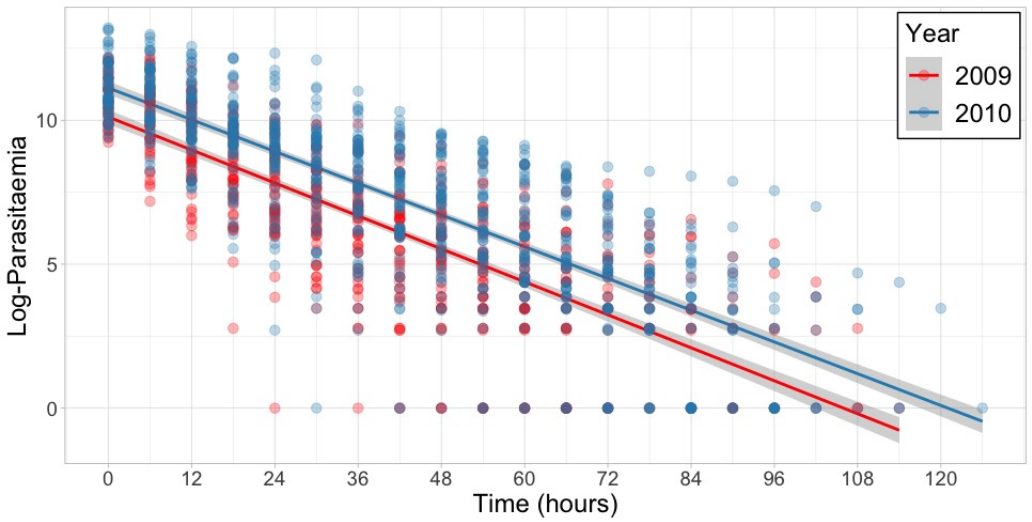

Figure 1: Log-Parasite counts (y_i) of all 110 patients from the Pursat data set, with separate regression lines and 95% confidence intervals for a linear model with year as a factor (Equation 1)

.

Approach 1: Linear Modelling

We can fit a basic linear regression model to the combined data set, treating each covariate as a factor. For example year with 2 levels; 2009 and 2010, as in Figure 1. In this model year_i has a value of 0 for data from 2009 and 1 for 2010, therefore the ith indexed data point is represented by the equation:

$$\log(y_i) = \alpha - \beta t_i + 1.19 \times \text{year}_i + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma^2) \qquad \textbf{(Equation 1)}$$

Performing a basic linear model analysis in R, the year term was evaluated as significant ($p << 0.01$) but an interaction term ($\text{year}_i \times t_i$) was not ($p > 0.1$) and is therefore excluded from the model. This suggests there is a difference in initial parasitaemia between years, but not rate of clearance, when no other covariates are considered.

The advantage of this method is its simplicity and the well developed theory around linear models and their analysis. However - by treating all data points equally we are unable to represent the connection between measurements from the same individual. This method is therefore unsuitable as it is unable to estimate the between subject variation in treatment effects.

Approach 2: Proportion of Patients with Detectable infection

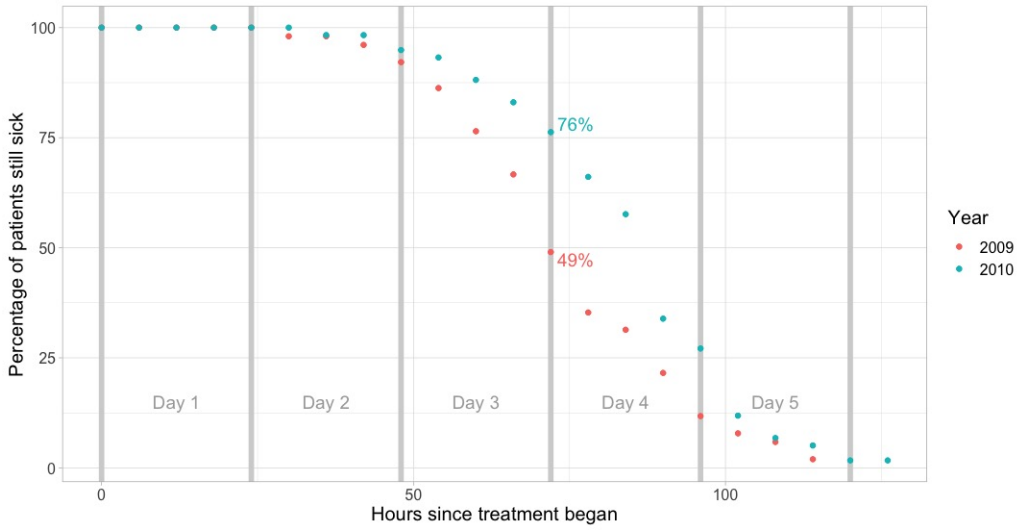

Figure 2: Percentage of patients with detectable malaria infections at the end of each treatment day, separated by year.

Many studies use the proportion of patients with any detectable infection on each day of treatment, or the time taken until parasite density levels are below the detection limit (Figure 2) (Stepniewska et al 2010, Dahal et al 2015). This method is very easy to implement and can be calculated without any regular intervals

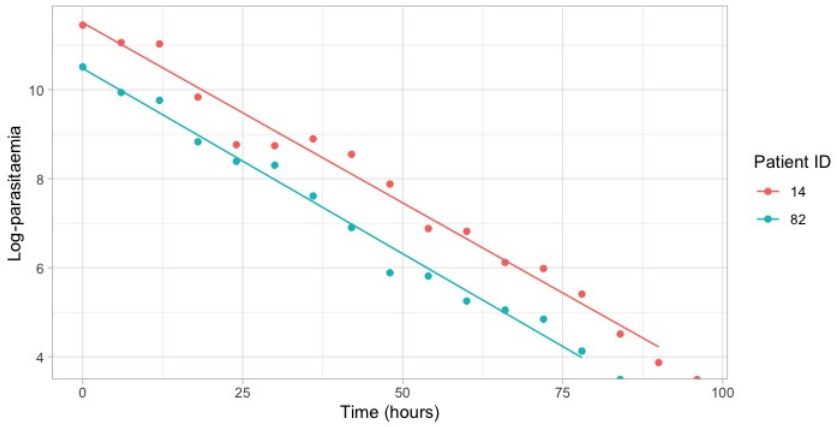

Figure 3: Parasite clearance profiles from patients 82 and 14 with linear regression lines to demonstrate similarity of the slopes.

of blood sampling required. The significant disadvantage however is that duration of infection has a strong correlation with initial parasitaemia. Even with near-identical slopes of parasite clearance, an infection with a higher density of initial parasites will take longer to clear (Figure 3).

Another concern is the reduced reliability of Microscopy measurements at low parasite densities (O'Meara et al 2005), one study (Berzosa et al 2018) found 19.4% of blood samples assessed as clear of malaria were false-negatives when reassessed with PCR. Alternatively using PCR diagnostics can greatly increase measurement accuracy. Therefore percentage of infected patients is at risk of being a poor indicator of treatment effect.

Approach 3: Parasite Clearance Rate

One of the most robust measures of treatment effectiveness is the declining log-linear slope of parasite density (Ashley et al 2014, Flegg et al 2011, White 2011). However creating a reliable and accurate method of finding this slope is complicated by a few factors. Parasite clearance measurements may exhibit 'lag' and 'tail' phases. The prevalence of these has been estimated as approximated 30% (Flegg et al 2011). The 'lag' phase occurs when the parasite readings remain constant (or even increase by a small degree) after treatment is administered. This phase is then followed by a period of log-linear decay, and finally a 'tail' with repeated measurements near or at 0. These phases occur due a variety of biological and methodological factors and are not considered representative of treatment effect. Therefore identifying and removing these phases from calculation of parasite clearance slope will increase accuracy of the results.

Parasite Clearance Estimator

The Parasite Clearance Estimator is a tool developed for the Worldwide Antimalarial Resistance Network (WWRAN) (Flegg et al 2011). This standardised system for calculating slope (or equivalently half-life) of *P. falciparum* infections is easily available as an online tool. The methods include consideration of 'lag' and 'tail' phases, as well as a purpose-made outlier detection tool to identify and censor biologically improbable or

impossible measurements.

This system has become widely used throughout malaria research. For example 68% (17 of 25) of papers with suitable data between Nov 2011 and May 2014 used this tool (Fogarty et al 2015). Therefore the PCE represents a significant improvement on previous methods - particularly in terms of modelling the possible presence of 'lag' and 'tail' phases, and being generally quite accessible for researchers.

Despite its usage, there are some flaws in how the PCE tool is utilised in practice. Generally, analysis is done in 2 stages: **(1)** parasite half-lives are calculated using the PCE, and **(2)** regression is performed on these summary statistics, in order to assess the possible impacts of covariates. As an example of the possible ways to look at these summary statistics, Figure 4 shows the slope coefficients from patients in the Pursat data set, separated by year.

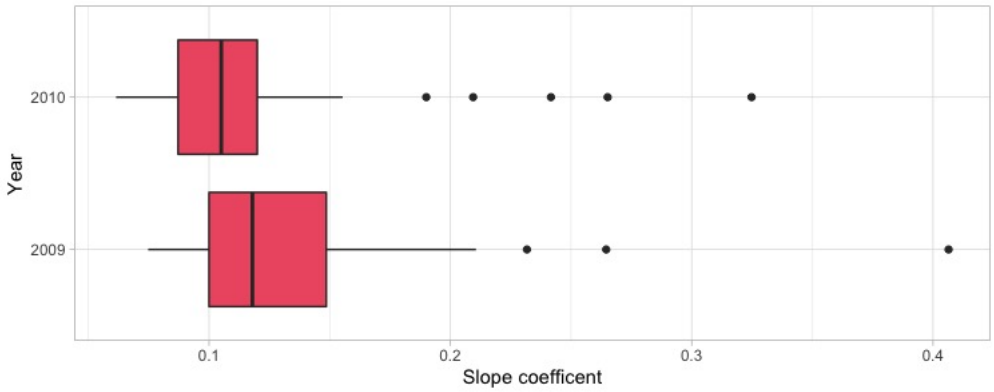

Figure 4: Slope coefficients for log-linear parasite decay rate by year, calculated using the Parasite Clearance Estimator tool.

A weakness of this approach is that half-life values are being treated as primary data, and therefore during regression they are assumed to be *homoscedastic* (have homogeneous error or noise values). This assumption however, does not hold for half-life estimates calculated with the PCE as the number of readings from a patient can vary widely. For example there should be significantly less error expected for a slope calculated from a large number measurements, than one calculated from very few. By separating the analysis into these two distinct parts, we loose some information that could help to inform the covariate analysis.

Bayesian PCE

In 2015, Fogarty et al. proposed a Bayesian approach to modelling parasite clearance which would allow the 'two-stage approach' to become integrated. This method uses a hierarchical model, where each patient has their own individual slope, intercept and lag/tail change-point parameters. Each of these patient-level parameters are described as independent and identically distributed (iid) draws from a population-level distribution, described by population-level hyper-parameters. This is the method adapted for the modelling in this report.
The main advantages of this method is that unlike the PCE it allows for the inclusion of patients with less than 3 measurements (the systematic exclusion of which could possibly bias results). It also enables all information

from the slope analysis to be included in the estimation of the association between the individual slope parameters and covariates of interest

MODELLING METHODS:

A newer approach (as conceived by Fogarty et al in 2015) is to use a Bayesian hierarchical framework, this manages to capture the individual level variation, but includes a representation of the population level effects. This is the model adapted and implemented in this report. The following elements of relevant mathematical theory may be helpful for understanding how it works:

Bayesian Modelling

Bayesian statistics is based upon Bayes Theorem, $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, which describes probabilities in the context of previously occurring events. Information that is already known is represented as a *Prior Distribution*. Whereas the final or *Posterior Distribution* is a combination of this previous knowledge and the data we input. The more data we have, the less heavily the prior influences our model. The conditional formula above can be utilised in a modelling context where θ represents our parameters and y represents the observed data:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Hierarchical Modelling

A hierarchical model enables the modelling of more complex dependency structures. To introduce a hierarchy, we represent each individual's parameters (eg. slope and intercept) as iid draws from a joint distribution. In this way, 'global' (population level) parameters define the distributions from which 'local' (patient level) parameters are drawn. This separates out population level affects from individual affects and allows both to coexist together.

Sampling

Generally, the overall likelihood equation for a Bayesian model can be quite complex and difficult to write in a closed form. One solution is to implement sampling algorithms which simulate draws from the posterior distribution through a variety of methods. Summary statistics from these samples, like the mean, median and 95% credible intervals, are used to draw inference about this posterior distribution.

Stan

Stan is a software that can be interfaced through a variety of programs, for this project I used R to write and run my Stan code. Stan allows for sophisticated sampling techniques to be implemented without the user needing to laboriously code the sampling algorithm. The user only needs to specify the inputs and required output: data, parameters, model (likelihood) and any generated parameters such as posterior predictive samples.

MODELLING IMPLEMENTATION:

The sampling was run for 4 chains with 4000 iterations each, taking approximately 50 minutes to run. The first 500 iterations of each chain were discarded as burn-in, leaving 14 000 iterations total. Figure 5 shows 95% credible intervals from the posterior predictive distributions of the first 20 patients.

The adapted Bayesian hierarchical model (based on that proposed by Fogarty et al in 2015) for Parasite Clearance has the following likelihood for the ith patient at time j:

$$\log(y_{ij}) = \alpha_i - \beta_i(\delta_i^l \mathbb{I}(t_{ij} < \delta_i^l) + t_{ij}\mathbb{I}(\delta_i^l < t_{ij} < \delta_i^\tau) + \delta_i^\tau \mathbb{I}(t_{ij} > \delta_i^\tau)) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

The likelihood can be equivalently represented using cases:

$$\log(y_{ij}) = \begin{cases} \alpha_i - \beta_i \delta_i^l + \epsilon_{ij}, & t_{ij} < \delta_i^l \\ \alpha_i - \beta_i t_{ij} + \epsilon_{ij}, & \delta_i^l < t_{ij} < \delta_i^\tau \\ \alpha_i - \beta_i \delta_i^\tau + \epsilon_{ij}, & t_{ij} > \delta_i^\tau \end{cases}$$

The values δ_i^l and δ_i^τ represent non-negative change-points for the end of the 'lag' phase and beginning of the 'tail' phase respectively. The restriction for each patient i is that $\delta_i^l + \delta_i^\tau \leq \max(t_i j)$. $\mathbb{I}(\cdot)$ represents the indicator function, having a value of 1 when the given inequality is satisfied, and 0 otherwise.

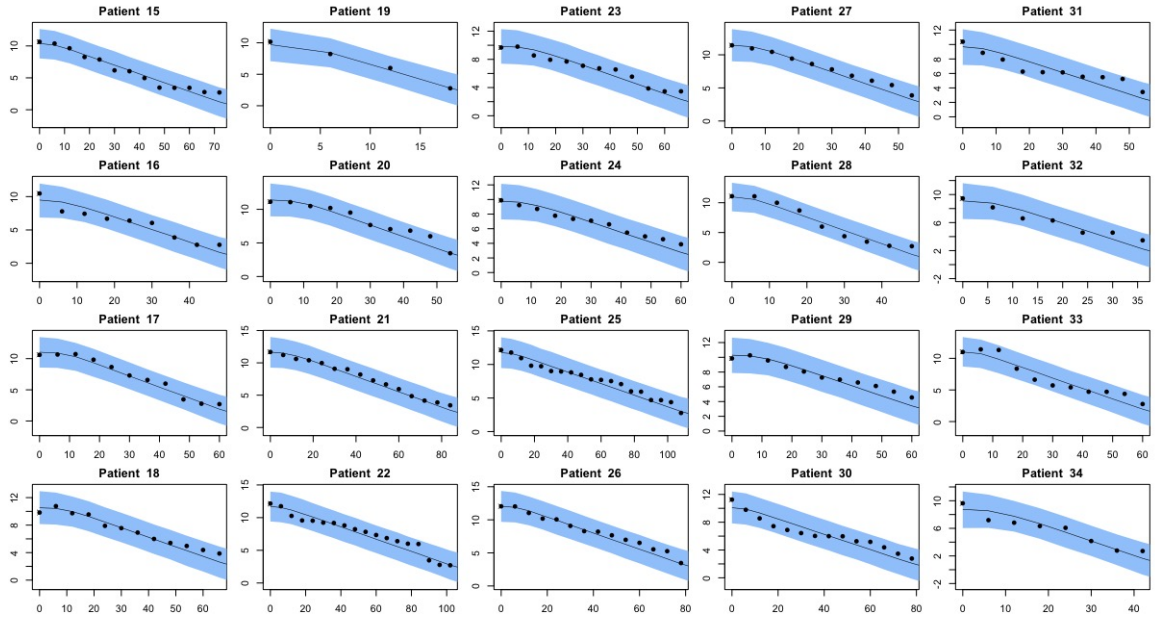

Figure 5: Log-parasite counts of patients 1 to 20 over time, with 95% posterior predictive credible intervals, for a lag-tail model. The x axes are time in hours and the y axes are log-parasitaemia densities. The median is in black.

The parameters for each patient are modelled as coming from these joint 'population level' distributions:

$$\log(\alpha_i) \sim N(\eta, \sigma_\alpha^2), \quad \log(\beta_i) \sim N(\gamma, \sigma_\beta^2)$$
$$\log(\delta_i^l) \sim N(2,1), \quad \log(\delta_i^\tau) \sim N(a,1), \quad a \sim \text{Uniform}(0,4)$$

With uniform priors on $\gamma, \eta, \sigma_\alpha^2, \sigma_\beta^2$ and σ_ϵ^2 : $p(\eta, \sigma_\beta^2) \propto 1, \quad p(\gamma, \sigma_\alpha^2) \propto 1, \quad p(\sigma_\epsilon^2) \propto 1$

The original model uses more complex priors:

$$\log(\delta_i^l) \sim N(a,c), \quad \log(\delta_i^\tau) \sim N(b,d), \quad a,b \sim \text{N}(\log(6), 0.5^2), \quad c,d \sim \text{InvGamma}(1,1)$$

As well as additional parameters to represent the overall probability of a lag or tail phase existing. These were excluded from the adapted model for the sake of simplicity.

Detection Limit and Censoring

Every patient has a final reading of 0 parasites and this presents 2 issues - firstly we cannot take the logarithm of 0, and secondly we don't actually know the value was 0 - we only know it was below an observable density for the methods used. In this way the data in censored, meaning a value lower than the detection limit cannot be recorded. For the Pursat data used here, the microscopist examined a blood sample of 500 white blood cells, recorded the number of parasite cells and multiplied this by 16 to calculate the number of parasites per 8000 white blood cells, (the assumed number present in a μL). This means 1 observed parasitised cell would be recorded as a density of 16 and 2 as 32 and so on.

There are a number of standard approaches to censored data sets. For this analysis I represented the likelihood equation in 2 parts, one normal and one from the normal cdf (Φ). This can be called the 'M3' method (Bergstrand Karlsson 2009) and can reduce bias compared to simply removing or replacing the censored values. The lowest non-zero count recorded in the data set is 15, therefore this is a suitable lower bound for the non-censored distribution.

$$\log(y_{ij}) = \begin{cases} \frac{1}{2\pi\sigma^2} \exp[\frac{-1}{2\sigma^2}(\log(y_{ij}) - \log(\hat{y}_{ij}))^2], & y_{ij} \geq 15 \\ \Phi\left(\frac{\log(15) - \log(\hat{y_{ij}})}{\sigma}\right) & y_{ij} = 0 \end{cases}$$

Where: $\hat{y}_{ij} = \alpha_i - \beta_i(\delta_i^l \mathbb{I}(t_{ij} < \delta_i^l) + t_{ij}\mathbb{I}(\delta_i^l < t_{ij} < \delta_i^\tau) + \delta_i^\tau \mathbb{I}(t_{ij} > \delta_i^\tau))$

Diagnostics and Evaluation

There are a number of diagnostic tools available to evaluate the performance of the sampler. Although absence of evidence is not evidence of absence, it's a helpful indication that nothing is obviously incorrect. Table 1 presents the mean values of the population-level parameters, along with their calculated Effective Sample Size (ESS) and R-Hat. Effective Sample Size (ESS) represents an approximation of the amount of information the sampling contains, because draws are not completely independent we expect this to be lower than the sample size, but not too small as to be concerning. R-Hat is the ratio of between-chain residual error

Parameter	Mean	2.5%	97.5%	ESS	R-Hat
η	14.88	14.59	15.33	2546	1.00
σ^2_α	0.14	0.11	0.16	3466	1.00
γ	0.27	0.20	0.28	1141	1.00
σ^2_β	0.32	0.28	0.37	2410	1.00
σ^2_ϵ	1.21	1.17	1.24	14406	1.00

Table 1: Posterior Means and 95% credible intervals for key parameters, as well as the ESS (Effective Sample Size) and R-Hat values.

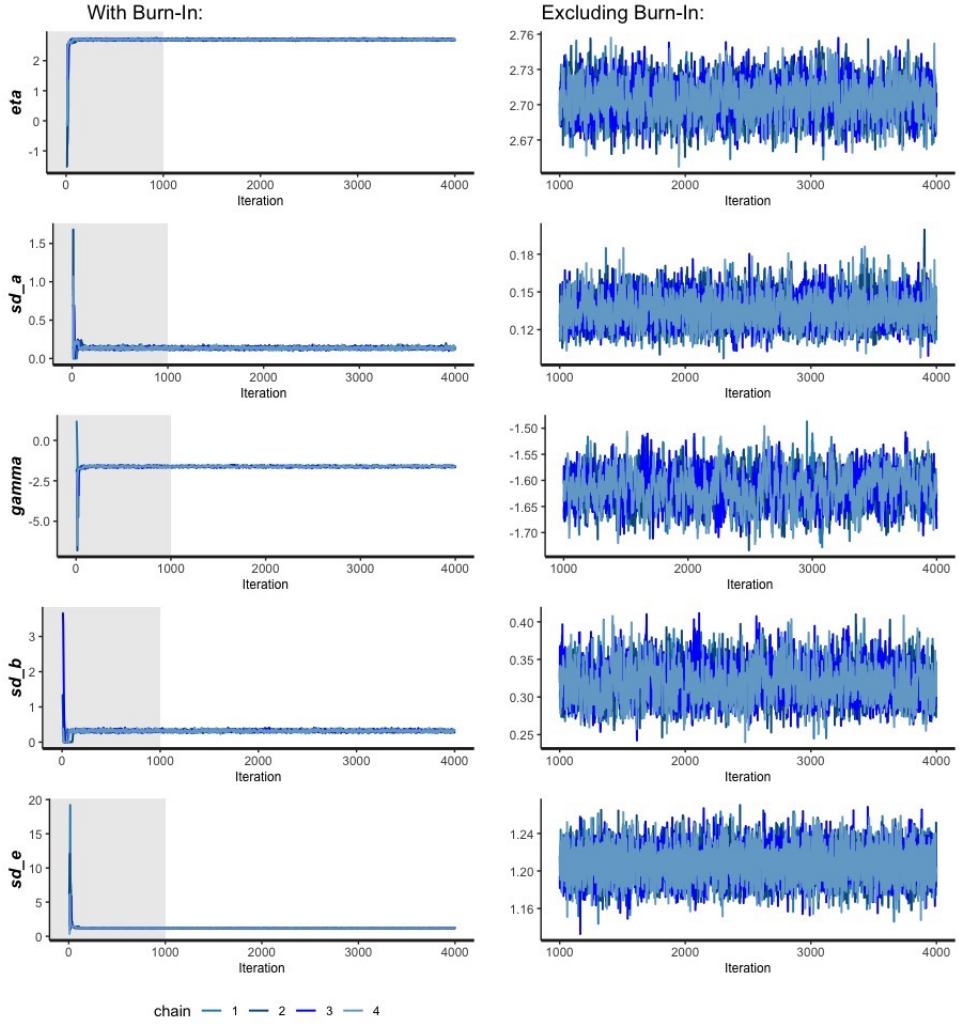

Figure 6: Trace plots for 4 chains of sampling with and without burn-in iterations, note γ and η are sampled on the log scale.

and within-chain residual error, if this value is significantly higher than 1 that would be cause for concern because it indicates the chains do not behave similarly. The ESS values seems sufficiently high, and the R-Hat

values sufficiently close to 1.

Figure 6 presents the trace-plots of these same parameters and demonstrates that with a range of starting values all the chains converge, mix well and behave similarly.

Comparison with Existing Methods

The Gibbs sampler utilised by Fogarty et al in 2015 took 32 hours to process in 2015. In comparison the methods presented here ran in approximately 50 minutes on a computer (itself purchased in 2016). It's worth noting that in 2019 an R package 'bhrcr' was released (Sharifi-Malvajerdi et al 2019) which reproduced the methods in Fogarty et al (2015). This package should increase accessibility of the Bayesian hierarchical model, however as of late 2020 it is no longer compatible with the latest version of R and is yet to be utilised for published data analysis.

Figure 7 compares the distributions of 110 patients log-linear slope values across 3 methods, the PCE, the Gibbs sampler created by Fogarty et al in 2015 and the methods developed here in Stan. The slopes are clearly comparable but not entirely similar to those calculated previously. Therefore this may be evidence of a weakness in the new Stan methodology.

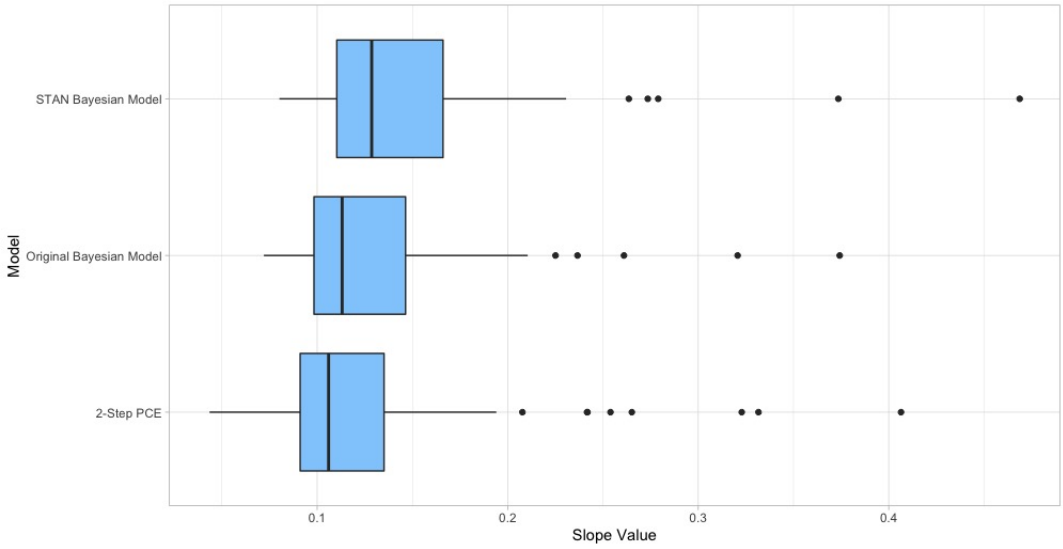

Figure 7: Comparison of the distribution of log-linear slope values (on the log scale) for the Pursat data set, calculated using the Parasite Clearance Estimator (PCE), a Gibbs sampler for the Bayesian Hierarchical methods and a Stan sampler for an adaptation of the Bayesian framework.

Introducing Covariate Analysis

The next goal of the analysis is to measure the impact of relevant covariates. This is done by introducing a matrix of covariate values $\mathbf{X}$ and altering $\boldsymbol{\gamma}$ to become a vector of length one more than the number of variables.

$$\log(\beta_i) \sim N(\boldsymbol{\gamma}\mathbf{X}, \sigma_\beta^2)$$

In this way each element of the γ vector represents a variable and therefore it's impact on the slope.

This sampler ran in 50 minutes for again 4 chains and a total of 14 000 iterations, however it did not reproduce the variable estimates calculated using previous methods. Diagnostic values like ESS and Rhat suggested the chains converged and behaved well, which implies that the sampling was run successfully, however the altered likelihood may have weakened the model's effectiveness.

DISCUSSION AND CONCLUSIONS:

This report explored an alternative method for modelling the parasite clearance rate for *P. falciparum* using an R interface in Stan. The sampling procedure was based on that developed by Fogarty et al in 2015 and used a Bayesian Hierarchical framework to balance out the representation of population-level and individual-level effects. This approach should create more powerful covariate analysis by integrating the 2-stage approach, and allowing for the inclusion of patient profiles with less than 3 measurements recorded.

The sampler ran efficiently and captured some of the profile shapes, however it failed to created reliable parameter estimates. Future work would be adjusting the priors of the sampler in order to reliably recreate the covariate estimates, as well as testing the analysis on a different data set.

Aside from inconsistent variable effect estimates, the Stan model is different from existing methods in terms of the statistical knowledge and programming skills required to implement suitable Bayesian sampling. This format is significantly more accessible than coding the sampler from scratch, and it is also easily adjustable and customisable, allowing for the methods to be updated to try new approaches or if new information becomes available. Therefore this report has successfully demonstrated the usefulness of the Stan software for increasing the use of more advanced Bayesian modelling techniques.

Acknowledgements:

Thank you to my supervisors A/Prof Jennifer Flegg and Dr Sophie Zaloumis for their considerable support, patience and advice during this project.

References:

Amaratunga, C, Sreng, S, Suon, S, Phelps, ES, Stepniewska, K, Lim, P, Zhou, C, Mao, S, Anderson, JM, Lindegardh, N, Jiang, H, Song, J, Su, X zhuan, White, NJ, Dondorp, AM, Anderson, TJC, Fay, MP, Mu, J, Duong, S & Fairhurst, RM 2012, 'Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: A parasite clearance rate study', *The Lancet Infectious Diseases*, vol. 12, no. 11, pp. 851–858.

Bergstrand, M & Karlsson, MO 2009, 'Handling data below the limit of quantification in mixed effect models', *AAPS Journal*, vol. 11, no. 2, pp. 371–380.

Dahal, P, d'Alessandro, U, Dorsey, G, Guerin, PJ, Nsanzabana, C, Price, RN, Sibley, CH, Stepniewska, K & Talisuna, AO 2015, 'Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data', *BMC medicine*, vol. 13, p. 212.

Dondorp, AM, Nosten, F, Yi, P, Das, D, Phyo, AP, Tarning, J, Lwin, KM, Ariey, F, Hanpithakpong, W, Lee, SJ, Ringwald, P, Silamut, K, Imwong, M, Chotivanich, K, Lim, P, Herdman, T, An, SS, Yeung, S, Singhasivanon, P, Day, NPJ, Lindegardh, N, Socheat, D & White, NJ 2009, ' Artemisinin Resistance in Plasmodium falciparum Malaria', *New England Journal of Medicine*, vol. 361, no. 5, pp. 455–467.

Flegg, JA, Guerin, PJ, White, NJ & Stepniewska, K 2011, 'Standardizing the measurement of parasite clearance in falciparum malaria: The parasite clearance estimator', *Malaria Journal*, vol. 10.

Fogarty, CB, Fay, MP, Flegg, JA, Stepniewska, K, Fairhurst, RM & Small, DS 2015, 'Bayesian hierarchical regression on clearance rates in the presence of "lag" and "tail" phases with an application to malaria parasites', *Biometrics*, vol. 71, no. 3, pp. 751–759.

Hamilton, WL, Amato, R, van der Pluijm, RW, Jacob, CG, Quang, HH, Thuy-Nhien, NT, Hien, TT, Hongvanthong, B, Chindavongsa, K, Mayxay, M, Huy, R, Leang, R, Huch, C, Dysoley, L, Amaratunga, C, Suon, S, Fairhurst, RM, Tripura, R, Peto, TJ, Sovann, Y, Jittamala, P, Hanboonkunupakarn, B, Pukrittayakamee, S, Chau, NH, Imwong, M, Dhorda, M, Vongpromek, R, Chan, XHS, Maude, RJ, Pearson, RD, Nguyen, T, Rockett, K, Drury, E, Gonçalves, S, White, NJ, Day, NP, Kwiatkowski, DP, Dondorp, AM & Miotto, O 2019, 'Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study', *The Lancet Infectious Diseases*, vol. 19, no. 9.

Imwong, M, Suwannasin, K, Kunasol, C, Sutawong, K, Mayxay, M, Rekol, H, Smithuis, FM, Hlaing, TM, Tun,

KM, van der Pluijm, RW, Tripura, R, Miotto, O, Menard, D, Dhorda, M, Day, NPJ, White, NJ & Dondorp, AM 2017, 'The spread of artemisinin-resistant Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational study', *The Lancet Infectious Diseases*, vol. 17, no. 5.

Menard, D & Dondorp, A 2017, 'Antimalarial Drug Resistance: A Threat to Malaria Elimination', *Cold Spring Harbor Perspectives in Medicine*, vol. 7, no. 7.

O'Meara, WP, Permpanich, B, Wongsrichanalai, C, Forney, JR, Magill, AJ, Gasser, RA, Mckenzie, FE & Lucas, C 2005, 'Sources of Variability in Determining Malaria Parasite Density by Microscopy', *The American Journal of Tropical Medicine and Hygiene*, vol. 73, no. 3.

Ouji, M, Augereau, JM, Paloque, L & Benoit-Vical, F 2018, 'Plasmodium falciparum resistance to artemisinin-based combination therapies: A sword of Damocles in the path toward malaria elimination', *Parasite* vol. 25.

R Core Team. R 2020, 'A Language and Environment for Statistical Computing', *R Found Stat Comput*, Ver 4.0.3.

Sharifi-Malvajerdi, S, Zhu, F, Fogarty, CB, Fay, MP, Fairhurst, RM, Flegg, JA, Stepniewska, K & Small, DS 2019, 'Malaria parasite clearance rate regression: an R software package for a Bayesian hierarchical regression model', *Malaria Journal*, vol. 18, no. 1.

Stepniewska, K, Ashley, E, Lee, SJ, Anstey, N, Barnes, KI, Binh, TQ, D'Alessandro, U, Day, NPJ, De Vries, PJ, Dorsey, G, Guthmann, JP, Mayxay, M, Newton, PN, Olliaro, P, Osorio, L, Price, RN, Rowland, M, Smithuis, F, Taylor, WRJ, Nosten, F & White, NJ 2010, 'In vivo parasitological measures of artemisinin susceptibility', *Journal of Infectious Diseases*, vol. 201, no. 4, pp. 570–579.

Stan Development Team 2020, 'Stan Modeling Language Users Guide and Reference Manual', , no. 2.21, accessed from https://mc-stan.org.

Ashley, EA, Dhorda, M, Fairhurst, RM, Amaratunga, C, Lim, P, Suon, S, Sreng, S, Anderson, JM, Mao, S, Sam, B, Sopha, C, Chuor, CM, Nguon, C, Sovannaroth, S, Pukrittayakamee, S, Jittamala, P, Chotivanich, K, Chutasmit, K, Suchatsoonthorn, C, Runcharoen, R, Hien, TT, Thuy-Nhien, NT, Thanh, NV, Phu, NH, Htut, Y, Han, K-T, Aye, KH, Mokuolu, OA, Olaosebikan, RR, Folaranmi, OO, Mayxay, M, Khanthavong, M, Hongvanthong, B, Newton, PN, Onyamboko, MA, Fanello, CI, Tshefu, AK, Mishra, N, Valecha, N, Phyo, AP, Nosten, F, Yi, P, Tripura, R, Borrmann, S, Bashraheil, M, Peshu, J, Faiz, MA, Ghose, A, Hossain, MA, Samad, R, Rahman, MR, Hasan, MM, Islam, A, Miotto, O, Amato, R, MacInnis, B, Stalker, J,

Kwiatkowski, DP, Bozdech, Z, Jeeyapant, A, Cheah, PY, Sakulthaew, T, Chalk, J, Intharabut, B, Silamut, K, Lee, SJ, Vihokhern, B, Kunasol, C, Imwong, M, Tarning, J, Taylor, WJ, Yeung, S, Woodrow, CJ, Flegg, JA, Das, D, Smith, J, Venkatesan, M, Plowe, C V., Stepniewska, K, Guerin, PJ, Dondorp, AM, Day, NP & White, NJ 2014, 'Spread of Artemisinin Resistance in Plasmodium falciparum Malaria', *New England Journal of Medicine*, vol. 371, no. 5, pp. 411–423.

White, NJ 2011, 'The parasite clearance curve', *Malaria Journal*, vol. 10.