2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Statistical analysis of multiplex serological data: new tools to evaluate malaria transmission in pre-elimination settings.

Promotor :
Prof. dr. Niel HENS

Promotor :
Dr. VINCENT SLUYDTS

Thao Le Thi Phuong

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Statistical analysis of multiplex serological data: new tools to evaluate malaria transmission in pre-elimination settings.

Promotor :
Prof. dr. Niel HENS

Promotor :
Dr. VINCENT SLUYDTS

Thao Le Thi Phuong
Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Contents

1 Introduction **1**

2 Methodology **5**

2.1 Univariate frailty model 6

2.2 Bivariate frailty model 8

2.3 Baseline force of infection 9

2.4 Data description 10

3 Result **13**

3.1 Exploratory Data Analysis 13

3.2 Univariate 15

3.2.1 SIR 15

3.2.2 SIRS 16

3.3 Bivariate 17

4 Discussion **21**

5 Conclusion **23**

A Appendix

B R code

List of Figures

1 Life cycle of Malaria parasite in human body. 1
2 Schematic representation of SIR and SIRS models 5
3 Planning of the surveys of the randomized community based trial 10
4 Profiles of log transform serological data 13
5 Profiles of log transform serological data: GLURP and CSP 14
6 Univariate SIR model and goodness of fit 15
7 Univariate SIRS model and goodness of fit 16
8 Bivariate shared SIRS-SIR model 18
9 Bivariate shared gamma frailty model: Goodness of fit 19
10 Bivariate shared gamma frailty model: marginal Force of infection 19

List of Tables

1 Definition of model fitted to serological data 10
2 Parameter estimates for SIR and SIRS univariate frailty gamma model 17
3 Bivariate shared gamma frailty model with treatment effect: Parameter estimates 20

1 Introduction

Malaria is a serious and life-threatening disease caused by a parasite called *Plasmodium*, which is transmitted through bites of infected female *Anopheles* mosquitoes (World Health Organization, 2013). According to the World Health Organization (WHO), in 2013, there were 97 countries and territories with ongoing malaria transmission, and 7 countries in the prevention of reintroduction phase, making a total of 104 countries and territories in which malaria is presently considered endemic. The number of susceptible people to malaria is approximately 3.4 billion globally (World Health Organization, 2013). Thus, malaria has remained one of the world leading tropical health concerns.

Five different *Plasmodium* species are known to cause a malaria infection in humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi*. The majority of disease morbidity due to malaria is caused by *P. falciparum* and *P. vivax*. While *P. falciparum* predominates in Africa, *P. vivax* infected cases have been found more frequently in Asia, Latin America, and some areas of Africa. Regarding the severity and seriousness of malaria progression, although the mortality rates are higher for *P. falciparum*, *P. vivax* has been considered to be more difficult to eradicate (Kondrashin et al., 2014). One of the reasons is that the *P. vivax* parasite exhibits a (dormant) liver stage (called hypnozoite), which enables them to reactivate and cause relapse in patients who already recovered from the first episode of illness (World Health Organization, 2013).

The life cycle of malaria parasites can be described in Figure 1[1]. When a female infected mosquito takes a human blood meal, malaria parasites in the mosquito's salivary gland (known as sporozoites) inoculate into the human body. They immediately head through the blood vessels to the liver. Once inside the liver, they start to grow, multiply into merozoites, destroy the liver cells, and invade red blood cells. The parasites mature in the red blood cells, continue to reproduce merozoites, then burst out of the cells and continue the process by infecting new red blood cells (Marcus, 2009). It is the blood stage of parasite that causes the malaria symptoms (Centre for disease control and prevention, 2014).

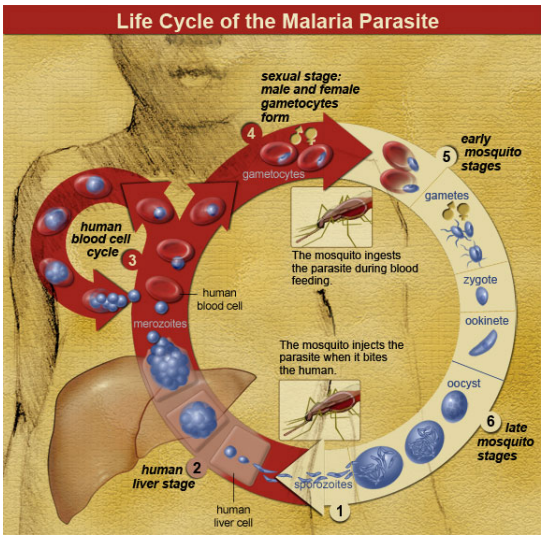


Figure 1: Life cycle of Malaria parasite in human body.

[1] source:http://www.niaid.nih.gov/topics/malaria/pages/lifecycle.aspx

In order to determine the disease burden, it is essential to monitor the transmission intensity of malaria. Although traditional methods such as the entomological inoculation rate (EIR: number of infectious mosquito bites per person per year) or parasite rate (PR) are widely used, they do suffer some drawbacks especially in low endemic areas. Because of the low frequency of positive samples, these methods often require a huge population sample to get an accurate estimate of malaria transmission. Furthermore, the heterogeneity in the mosquito distribution will as well require long term intensive sampling, and both entomological and parasitological measures are affected by seasonal fluctuation (Cook J. 2010). Thanks to the better understanding of malaria trends and prompt intervention policies, according to the latest WHO malaria estimates, malaria incidence reduced with more than 75% between 2000 and 2012 in 8 out of 10 countries in South East Asia. This situation has necessitated a new tool apart from traditional methods for accessing malaria transmission at low endemic levels in order to further the elimination process. In that context, serological tools can provide an alternative measure enabling to calculate disease transmission with higher efficacy compared to traditional methods. The human body keeps producing anti-malaria antibodies continuously for many months and even years after the antigen exposure. Hence, the presence of antibodies can be taken as a marker of previous infection (Webster et al., 1992). Using serological data in modeling malaria transmission has proven to be less susceptible to seasonality, and more sensitive (Corran P. 2004). Additionally, the simplicity in sampling method and analyzing making this method becomes more adaptive in poor conditions.

Classical antibody level measurements by using enzyme-linked immunosorbent assay (ELISA) has long been a standard in sero-epidemiological research. Despite of its high sensitivity, this method requires large amount of sample volume, laborious, and can only test one antibody at one time (Fouda et al., 2006). On the other hands, multiplex bead assay (MBAA) has been ascribed as a better replacement of monoplex methods. MBAA has the ability to assay multiple analysis simultaneously in a smaller volume of sero-sample, and provides time and cost-effective results (Ambrosino et al., 2010). Moreover, MBAA has proven to be easy to perform and yield similar sensitivity compared to ELISA assays (Elshal et al., 2006). For these reasons, MBAA become increasingly commonplace. Regarding the statistical context, by exploiting the results of more than one antigen per sample, we can combine more information to produce a more accurate estimate of malaria transmission.

In malaria context, serological data has been used in many previous studies to assess the epidemiology of the disease. In most of these studies, antibody levels are dichotomized into seropositive and seronegative status based on some cut-off points to obtain current status data. The model often fitted is the so called "catalytic" model, in which the disease prevalence is assumed to have an exponential distribution, implying a constant force of infection and sero-reversion rate for all age. Moreover, although using MBAA technique in testing and collecting data, most of studies that have been done so far only perform univariate analysis on antibody level without taken into account the multivariate nature of the data . Therefore, it is necessary to opt for a new statistical method that can take into account the association between two or more antibody levels. In the scope of the thesis, we study the association between two antibodies by ascribing dependency to individual heterogeneity other than age. To our knowledge, the idea was first introduced by Coutinho et al. (1999), applied in infectious context for the first time by Farrington et al. (2001) and further developed and refined by Hens et al. (2009) and Abrams et al. (2014).

The objectives of the thesis are:

- Develop a statistical method to incorporate unobserved heterogeneity in observed antibody response to malaria infection
- Extent the univariate model in a bivariate model taking into account two different antibody responses
- Adapt the model to be able to have covariate inside

The thesis is organized as follows. In the next section, we introduce different dynamic transmission models, followed by concepts about univariate and bivariate frailty model to capture individual heterogeneity. In Section 3, we describe multisera data on malaria taking from a large cross-sectional study in Cambodia in 2014. All the results are presented in Section 4. Lastly, we discuss the conclusion taken from the result part and end by proposing and suggesting some avenues for further research.

2 Methodology

Mathematical models have been used for decades to describe the process of disease transmission. In the context of this thesis, we consider the mathematical SIR and SIRS dynamic models under several assumptions. Firstly, we assume that the disease is at steady state which means the proportion of susceptible and infected individuals in the population are not changing over time. Secondly, we assume that the population has reached a demographic equilibrium or stationarity in age distribution which also implies constant birth (ie. the number of newborn child enter the population yearly). Finally, the death and birth rate are assumed to be constant over time, maintaining a balance population size N. Under these assumptions the transmission rates will be independent of calendar time, but depend on age only. Hence, this type of model is usually mentioned as time-homogeneous model. (Hens et al., 2012; Abrams et al., 2014)

In both the SIR and SIRS model, individuals of age a are assumed to be born into the susceptible (S) class. After an individual acquires an infection they transfer into the infection class (I) with rate $\lambda(a)$, the so called force of infection. It is the rate at which an individual acquires the disease, and generally assumed to be dependent on age and time ($\lambda(a,t)$). However, under time-homogeneous model, all parameters are constant with respect to time, therefore the dependence on two dimensions of time of the force of infection is reduced to age only: $\lambda(a,t) = \lambda(a)$.

In the SIR model, individuals acquire lifelong immunity and they permanently move into the recovered class (R) and then leave the transmission process. Whereas, for the SIRS model, individuals do not attain permanent immunity, but temporarily recover at a rate γ, and become susceptible again at a replenishment rate σ. The mortality rates at all states are neglected. A schematic representation of the SIR and SIRS model is presented in Figure 2.

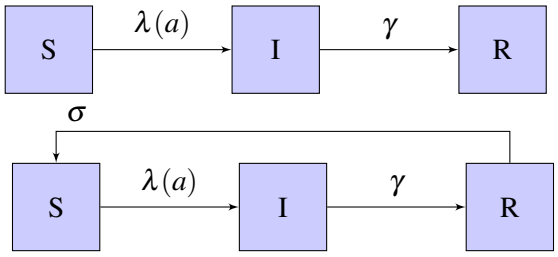


Figure 2: Schematic representation of SIR and SIRS models

The dynamics of the SIRS model can be described using following system of ordinary differential equations (ODEs):

$$\begin{aligned}
\frac{dS(a)}{dt} &= -\lambda(a)S(a) + \sigma R(a) \\
\frac{dI(a)}{dt} &= \lambda(a)S(a) - \gamma I(a) \\
\frac{dR(a)}{dt} &= \gamma I(a) - \sigma R(a)
\end{aligned}$$

where $S(a), I(a), R(a)$ represent the proportion of susceptible, infectious and recovered individuals of

age a in the population, respectively. The dynamics of the SIR model can be obtained by setting the replenishment rate σ in the SIRS model equal to zero.

2.1 Univariate frailty model

So far, we have assumed time homogeneous models in which all parameters only depend on age. However, in reality, individuals are greatly dissimilar with regard to the acquisition of an infection. This individual heterogeneity should be taken into account in order to not underestimate the force of infection (Wienke, 2010). In case heterogeneity caused by observed factors (eg. age, gender,...), some explanatory covariates might be included in the analysis. However, when heterogeneity caused by unknown covariates, it is more difficult to capture. In order to deal with such circumstance, frailty models were introduced by Vaupel et al. (1979). The main idea is to capture unobserved heterogeneity by incorporating a latent individual-specific variable called frailty, also known as activity level, denoted by Z. Z is a random variable that follows some distributions with the variance representing the heterogeneity among individuals in acquiring a particular infection (Hens et al., 2012; Coutinho et al., 1999). For more details about frailty models and its interpretation in medical context, we refer to Vaupel et al. (1979) and Morley et al. (2002).

The frailty term enters the dynamic model via the age-dependent force of infection $\lambda(a,Z)$. The force of infection now is conditional on individual frailty. The corresponding conditional susceptible proportion (also called survival function) for the SIRS model is given by:

$$\frac{dS(a|Z)}{dt} = -\lambda(a,Z)S(a|Z) + \sigma R(a|Z) \tag{1}$$

In low-endemic situation, the proportion of infected individuals in the population is relatively small compared to susceptible and recovery class, therefore we can plug in the approximation $R(a|Z) \sim 1 - S(a|Z)$ in equation 1 to obtain the marginal susceptible proportion (Abrams et al., 2014):

$$S(a|Z) = \exp\left(-\int_0^a \{\lambda(u,Z) + \sigma(u)\}du\right) + \int_0^a \sigma(u)\exp\left(-\int_u^a \{\lambda(v,Z) + \sigma(v)\}dv\right)du \tag{2}$$

Under the assumption of proportional hazards model (Cox et al., 1996), the random variable Z acts multiplicatively on the base line force of infection $\lambda_0(a)$, which means the frailty influences the force of infection in the same way at all ages:

$$\lambda(a,Z) = Z\lambda_0(a)$$

Consider the population level of survival function $S(a)$ as the mean of all individual survival functions with respect to Z. Therefore, the unconditional survival function can be obtained by integrating out the frailty term using a Laplace transform of the frailty distribution.

$$S(a) = E(S(a|Z))$$

$$S(a) = exp\left(-\int_0^a \sigma(u)du\right)\boldsymbol{L}(M_0(a)) + \int_0^a \sigma(u)exp\left(-\int_u^a \sigma(v)dv\right)\boldsymbol{L}(M_0(a) - M_0(u))du \quad (3)$$
$$= W_s(Q_\sigma)\boldsymbol{L}(M_0(a)) + \int_0^a \sigma(u)W_\sigma(Q_\sigma - Q_\sigma(u))\boldsymbol{L}(M_0(a) - M_0(u)du$$

where $W_\sigma(x) = \exp(-x)$, $Q_\sigma = \int_0^a \sigma(u)du$ is the cumulative replenishment rate and $M_0(a) = \int_0^a \lambda_0(u)du$ is the cumulative baseline hazard function, $\boldsymbol{L}(s)$ represents the Laplace transform of the random variable Z (Abrams et al., 2014).

As mentioned before, several distributions are feasible to characterize random variable Z. For the sake of simplicity in deriving unconditional survival function using Laplace transform, the Gamma distribution has been chosen for the frailty term based on its mathematical and computational applicability. To make sure the model is identifiable, we put a constraint on the frailty gamma distribution which results in mean of the distribution equal to 1 (Wienke, 2010).

$$Z \sim \Gamma(1/\sigma_f^2, 1/\sigma_f^2)$$

The variance of frailty variable σ_f^2 measures the heterogeneity across the population. Small σ_f^2 indicates that the frailty values are more concentrated around 1, whereas for large σ_f^2, the frailty values are more scattered and induce more heterogeneity in the force of infection, since $\lambda(a,Z) = Z\lambda_0(a)$. The corresponding Laplace transform for gamma frailty distribution takes the form $L(s) = (1+\sigma_f^2 s)^{-1/\sigma_f^2}$.

All the parameters estimates are obtained by maximizing the likelihood of serological data. Normally, after having test results of antibody levels, cut-off values are used to define disease status into sero-positive (contain specific antibody), sero-negative (not contain specific antibody) or equivocal (need further tests). However, setting a proper cut-off points is difficult to do when there is no clear separation in sero-status results, and thus can lead to information loss or bias estimates (Hens et al., 2012). Furthermore, if we could derive the prevalence from the distribution of antibody level, it is not necessary to use cut-off point to categorize each serum (Gay, 1996).

In order to use the antibody level directly, we turn into mixture models as a natural method to estimate the prevalence since sera samples are taken from a mixture of individuals who are infected and those who are not (Gay, 1996). In the univariate case, we assume antibody levels are coming from a two-component mixture distribution corresponding susceptible and infected subpopulations. The prevalence $\pi(a)$ is the proportion of seropositive, and hence the seronegative proportion is $1-\pi(a)$. $\pi(a)$ acts like a mixing probability or an age-dependent weighted parameter in the formula of likelihood which is specified as follows:

$$\ell = \sum(1-\pi(a))f_S(z_s|\boldsymbol{\theta_S}) + \pi(a)f_I(z_I|\boldsymbol{\theta_I})$$

where $z_j, (j = S, I)$ is antibody level of susceptible and infected individual deriving from density function $f_j(z_j|\boldsymbol{\theta_j})$.

2.2 Bivariate frailty model

Up to now, we described the application of an univariate frailty model as a way of dealing with heterogeneity among individuals due to some possible unobserved covariates. However, multiplex serological data enables us to study the association between multiple antibodies in acquisition of one or more infections. In order to achieve this, we would use bivariate gamma frailty model to capture the heterogeneity between two antibodies. The frailty model could be either shared or correlated.

The shared frailty indicates a common frailty term for both antibodies with respect to the acquisition of the disease. Denote Z_1, Z_2 as frailty terms for each antibody, then (Z_1, Z_2) follows a bivariate shared frailty gamma distribution which is characterized by a common frailty term Z (Hens et al., 2009; Hens et al., 2012; Farrington et al., 2001). In other words, we assume a perfect positive correlation among frailty terms, and conditional independence given the shared frailty between any two antibodies. The joint unconditional bivariate proportion of susceptible individuals can be derived as follows:

$$
\begin{aligned}
S_{12}(a) = {} & \boldsymbol{L}(M_{10}(a) + M_{20}(a)\{W_{\sigma_1}(Q_{\sigma_1}(a))W_{\sigma_2}(Q_{\sigma_2}(a))\} + \\
& \int_0^a \sigma_2(v) W_{\sigma_1}(Q_{\sigma_1}(a)) W_{\sigma_2}(Q_{\sigma_2}(a) - Q_{\sigma_2}(v)) \boldsymbol{L}(M_{10}(a) + M_{20}(a) - M_{20}(v)) dv + \\
& \int_0^a \sigma_1(u) W_{\sigma_2}(Q_{\sigma_2}(a)) W_{\sigma_1}(Q_{\sigma_1}(a) - Q_{\sigma_1}(u)) \boldsymbol{L}(M_{20}(a) + M_{10}(a) - M_{10}(v)) dv + \\
& \int_0^a \int_0^a \sigma_1(u)\sigma_2(v) W_{\sigma_1}(Q_{\sigma_1}(a) - Q_{\sigma_1}(u)) W_{\sigma_2}(Q_{\sigma_2}(a) - Q_{\sigma_2}(v)) \\
& \boldsymbol{L}(M_{10}(a) - M_{10}(u) + M_{20}(a) - M_{20}(v)) du dv
\end{aligned}
\tag{4}
$$

where σ_1 and σ_2 are replenishment rates for infection 1 and 2, M_{i0} is the cumulative baseline hazard function for antibody i, and $Q_{\sigma i}$ is the cumulative replenishment rate ($i = 1, 2$). Setting $\sigma_i(a) = 0$ will give us the SIR dynamic model for infection i.

The marginal survival function S_i ($i = 1, 2$) can be expressed as in the univariate case, i.e., formula 4 for SIRS infection, and the version for immunizing infection can be easily derived by setting the replenishment rate equal to 0.

A natural extension of shared frailty model is correlated frailty model, in which we allow a more flexible correlation between frailties. The frailty terms can be decomposed into 2 components: one is common for both frailties and one is specific to the specific infection (Hens et al., 2009).

$$
\begin{aligned}
Z_1 &= \sigma_{1f}^2 (Y_0^* + Y_1^*) \\
Z_2 &= \sigma_{2f}^2 (Y_0^* + Y_2^*) \\
Y_l^* &\sim \Gamma(k_l, 1), k = 0, 1, 2
\end{aligned}
$$

where σ_{if}^2 (i=1,2) represents the frailty variance. Because of the unit mean constraint on the frailty variable, the frailty variance $\sigma_{if}^2 = (k_0 + k_i)^{-1}$. The Pearson correlation among the frailty term equal $\rho_{12} = k_0 / \sqrt{(k_0 + k_1)(k_0 + k_2)}$. The additive structure implies an upper bound restriction on the correlation coefficient $0 \leq \rho \leq \min(\frac{\sigma_{1f}}{\sigma_{2f}}, \frac{\sigma_{2f}}{\sigma_{1f}})$. The correlated frailty model can be reduced to shared model when setting $Y_1^* = Y_2^*$. (Hens et al., 2009)

In order to construct the likelihood in bivariate case, we assume each individual can come from one of 4 different subpopulations: seropositive for one antibody and seronegative for another antibody, seropositive for both antibodies or seronegative for both antibodies. The 4 subpopulations are characterized by 4 corresponding density distributions: $f_{IS}, f_{SI}, f_{II}, f_{SS}$, of which, each density is a two-component mixture model with mean ($\boldsymbol{\mu}_{ij}$) and variance-covariance matrix $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{pmatrix}$ $(i, j = \{S, I\})$. The contribution of one individual in the bivariate likelihood function can be represented as:

$$\ell = p_{11}(a) f_{II}(z_1, z_2|\theta_{II}) + p_{10}(a) f_{IS}(z_1, z_2|\theta_{IS}) + p_{01}(a) f_{SI}(z_1, z_2|\theta_{SI}) + p_{00}(a) f_{SS}(z_1, z_2|\theta_{SS})$$

where z_1, z_2 are test result of two antibodies, and $p_{11}(a), p_{10}(a), p_{01}(a), p_{00}(a)$ define the multinomial probability distribution for antibody level given age. The multinomial probability can be expressed in term of marginal and joint survival function as follows:

$$\begin{aligned} p_{11}(a) &= 1 - S_1(a) - S_2(a) + S_{12}(a) \\ p_{10}(a) &= S_2(a) - S_{12}(a) \\ p_{01}(a) &= S_1(a) - S_{12}(a) \\ p_{00}(a) &= S_{12}(a) \end{aligned}$$

All the parameter estimates are then obtained by maximizing the log likelihood function.

2.3 Baseline force of infection

The base line force of infection is modelled using a parametric function. The advantage of having a parametric baseline force of infection is that it can assure that the marginal likelihood is fully parametric and can be maximized to obtain estimates for all parameters. We have opted for the Gompertz function.

$$\lambda_0(a) = \alpha \exp(\beta \cdot a)$$

In modeling infectious disease, exponential distribution is commonly used in a so called "catalytic model". This distribution implies a constant or age-independent force of infection. As can be seen, Gompertz distribution is more flexible, and can be reduced to exponential distribution by setting $\beta = 0$.

The relationship between baseline force of infection and unconditional force of infection can be expresses as follows (Wienke, 2010):

$$\lambda(a) = \frac{\alpha \exp(\beta \cdot a)}{1 + \sigma_f^2(\alpha/\beta)(\exp(\beta \cdot a) - 1)}$$

The force of infection can be extended by introducing an observed covariate (ie. treatment effect) into the function. Conditional on the frailty term and the explanatory covariate, the force of infection is of the form:

$$\lambda(a|X_{ij}, Z_i) = Z_i \lambda_0(a) \exp(\beta' X_{ij})$$

with X and β' are covariate and parameter. The unconditional force of infection can be obtained by integrating out the frailty variable, conditional on age and covariate X (Wienke, 2010):

$$\lambda(a|X) = \frac{\lambda_0(a)\exp(\beta' X)}{1+\sigma_f^2 M_0(a)\exp(\beta' X)}$$

In the context of the thesis, we consider several models adapting frailty terms in both univariate and bivariate cases for non-immunizing and life-long infections. The definition of models that have been fitted are described in Table 1. We use Akaike Information Criterion (AIC) value for model selection.

Table 1: Definition of model fitted to serological data

Model	Frailty	Dynamic	Replenishment rate σ
Model 1	Univariate	SIR	0
Model 2	Univariate	SIRS	σ
Model 3	Shared	SIRS-SIR	$\sigma_1, 0$
Model 4	Correlated	SIRS-SIR	$\sigma_1, 0$

The statistical analyses are performed using R software version 3.1.1 (R Core Team, 2014).

2.4 Data description

In 2012 and 2013, a study was conducted in Cambodia with the aim of evaluating an alternative malaria intervention: topical mosquito repellent in order to further eliminate the disease. It is a two arms study in which the control group was provided a large coverage of Long Lasting Insecticidal Nets (LLINs), whereas the intervention group had the same coverage of LLINs combined with the massive use of the topical mosquito repellent.

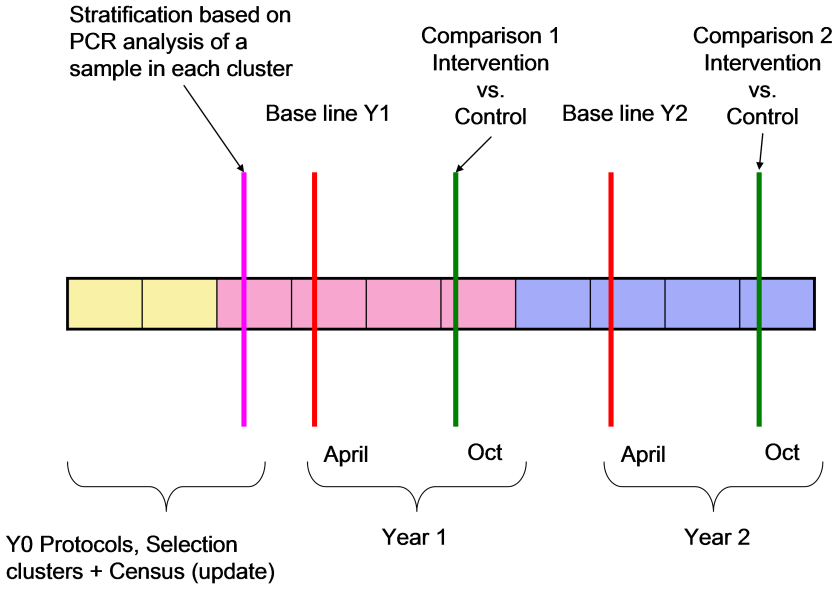


Figure 3: Planning of the surveys of the randomized community based trial

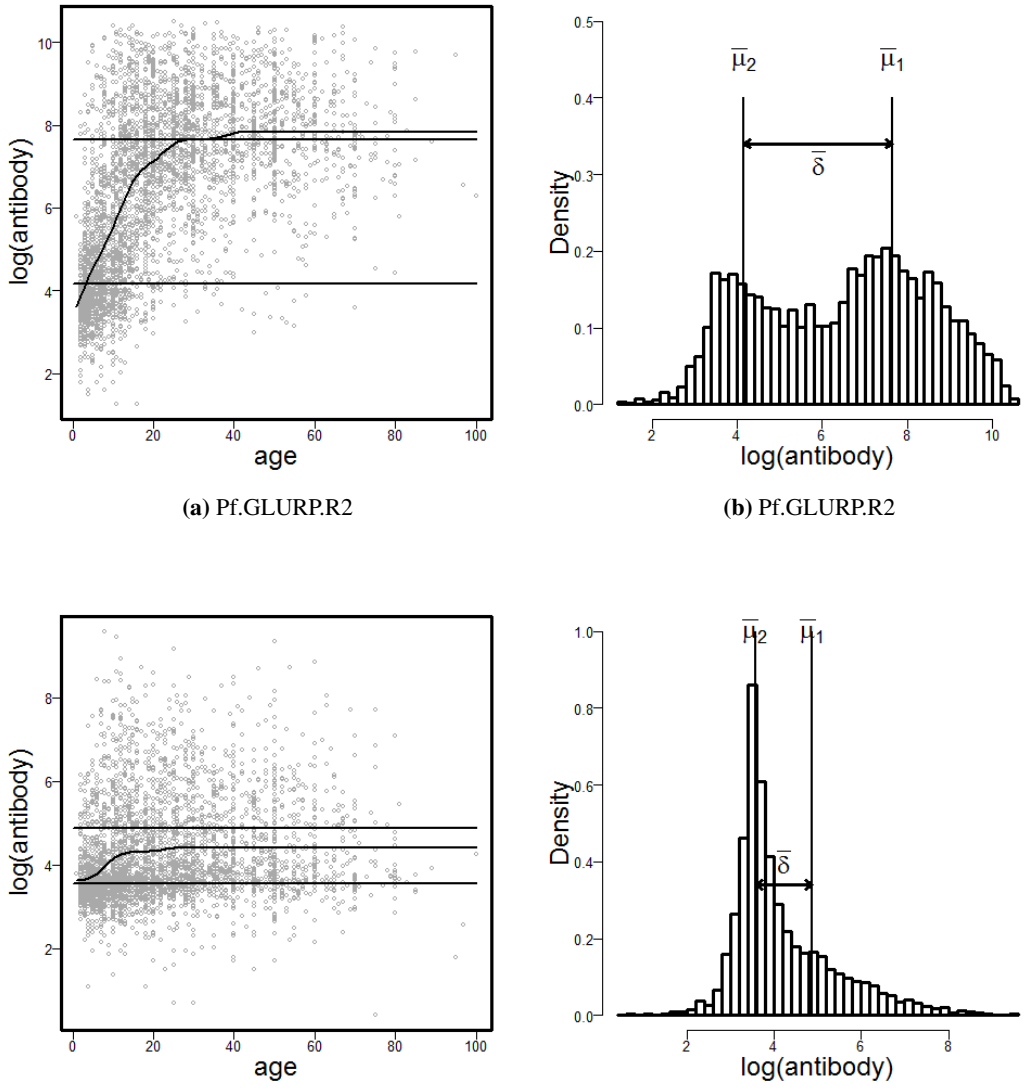
Sera sample were collected and tested for the presence of 21 antigens using multiplex bead assay

technique. In this study follow-up was carried out every year at two time points (Figure 3): at the beginning of rainy season (April), and six months after starting the intervention (October). A two year study is foreseen to tackle the annual variation in malaria transmission and to test for a cumulative effect from one year to another. The antibody levels data that directly used in following analysis were taken from survey two of the study.

3 Result

3.1 Exploratory Data Analysis

In multiplex technique, BSA (bovine serum albumin) are used as blocking and carrier protein, and can be considered as a control assay. After having the test results for all antibodies, we subtract them for the BSA values in order to correct for the background's noise. There are two typical profiles for data after correcting for background value and log transform as presented in Figure 4. As can be seen, the logarithm



(a) Pf.GLURP.R2 **(b)** Pf.GLURP.R2

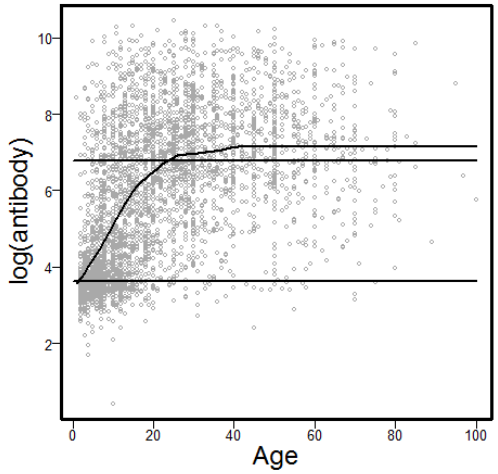
(c) LSA **(d)** LSA

Figure 4: Profiles of log transform serological data. Left panel: Scatter plot of log(antibody level), means for the two components, and the overall mean. Right panel: a histogram of the log(antibody level), $\bar{\delta}$ is the mean difference

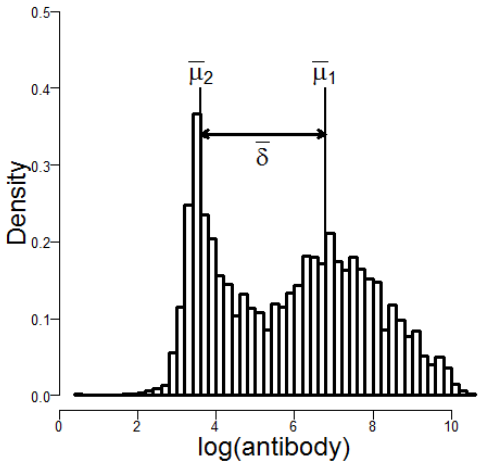
transform of antibody Pf.GLURP.R2 has a 2-component mixture distribution pattern with equal sample

size for the two serological status populations. Whereas, with LSA, we do not observe the same pattern. Instead, LSA has a long tail distribution, indicate that not many seropositive samples were collected.

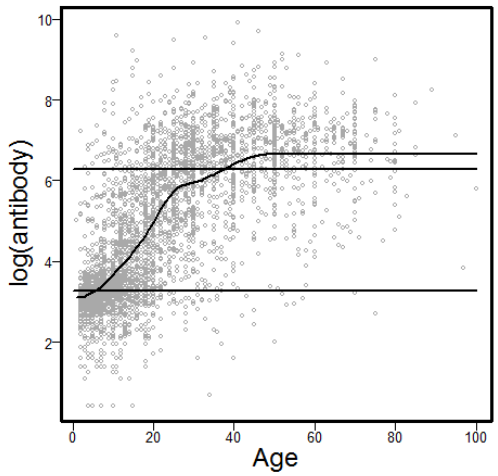
We illustrate the methodology on two antigens: GLURP and CSP whose histograms have bimodal curve. Profiles of others antigen are presented in the Appendix.



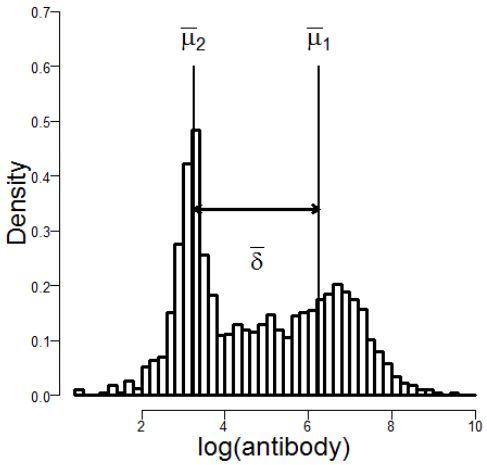
(a) GLURP



(b) GLURP



(c) CSP



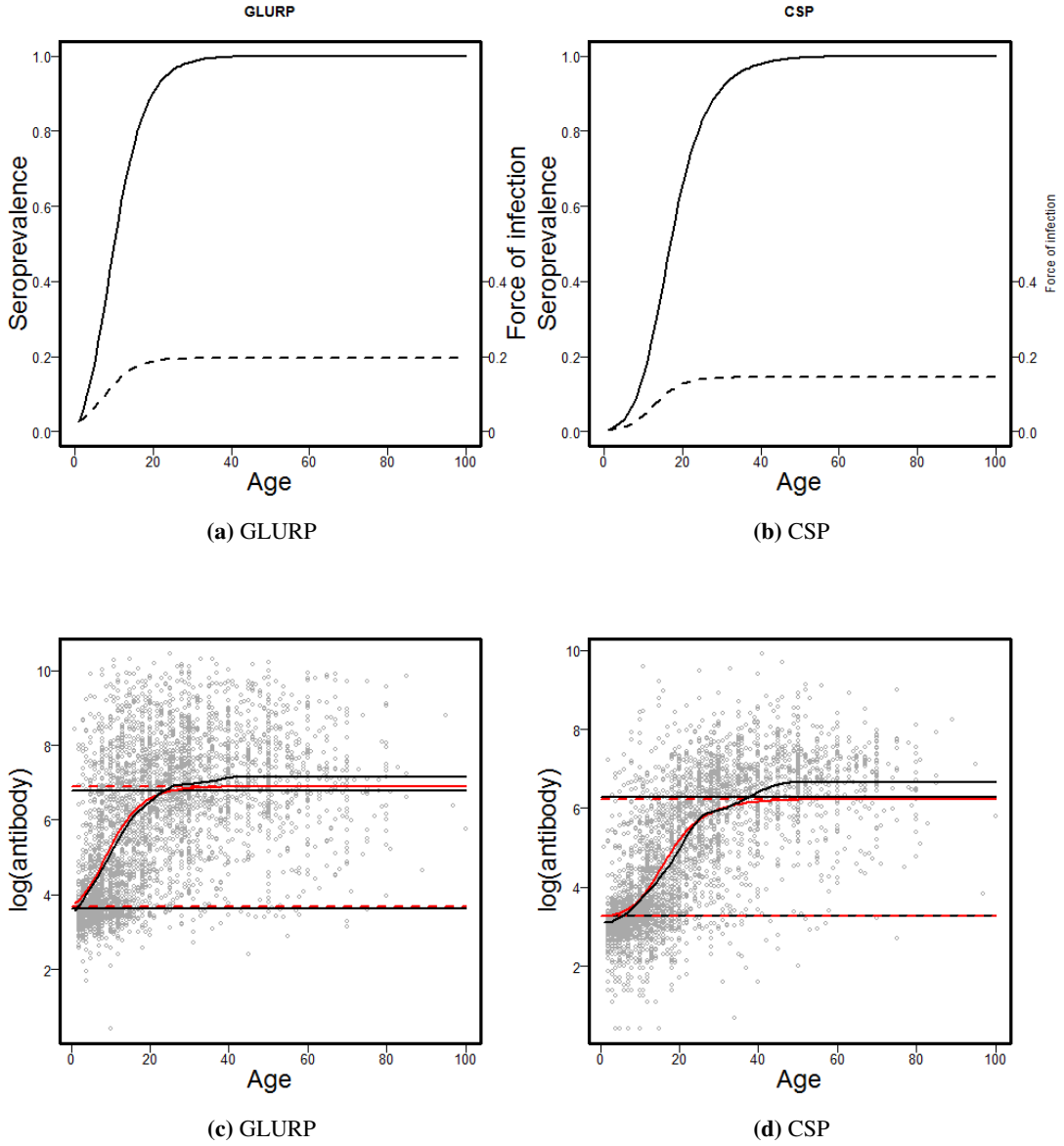
(d) CSP

Figure 5: Profiles of log transform serological data. Left panel: Scatter plot of log(antibody level), means for the two components and the overal mean. Right panel: a histogram of the log(antibody level), $\bar{\delta}$ is the mean difference

3.2 Univariate

3.2.1 SIR

Prevalence and force of infection obtained from SIR model for both antigens are presented in Figure 6. The force of infection for both antigens are increasing with age and remain constant from 20 years onward for GLURP and around 25 years onward for CSP. In general, CSP yields smaller FOI compared to GLURP.



(a) GLURP **(b)** CSP

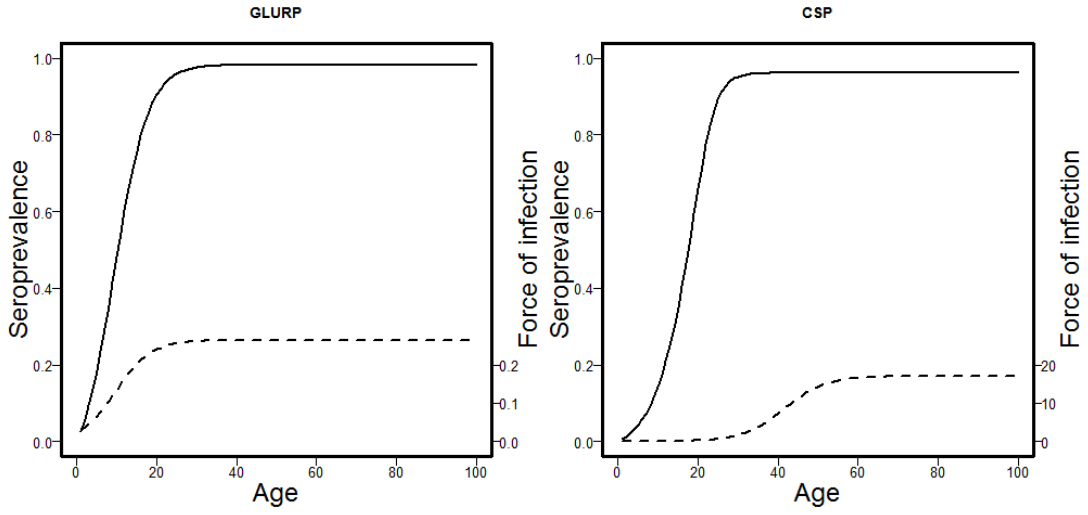
(c) GLURP **(d)** CSP

Figure 6: Univariate SIR model: Upper panel: seroprevalence estimate; Lower panel: Goodness of fit. Black solid lines: mean of two subpopulations estimated using mixdist function. Red dashed lines: estimate mean of two subpopulations with frailty model. Black curve: observed overall mean using spline method. Red curve: fitted overall mean with frailty model

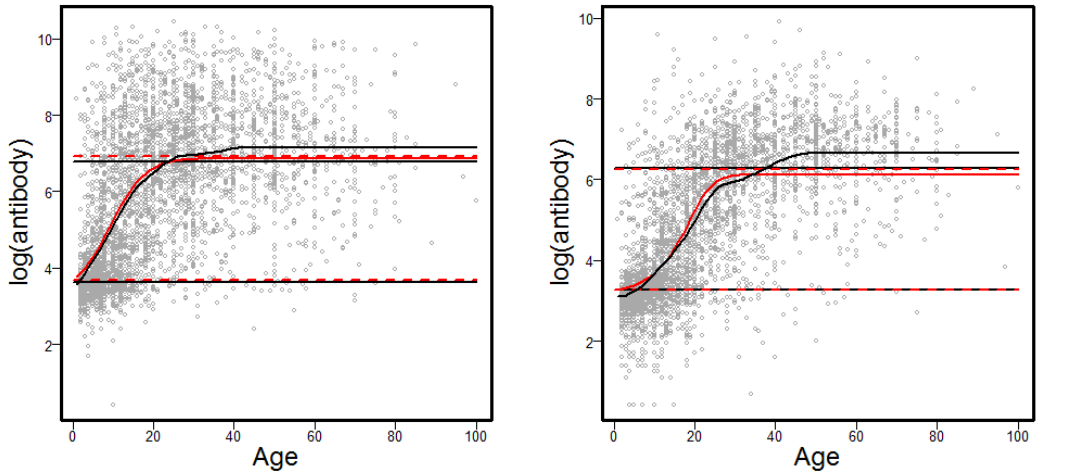
The goodness of fit of the models are visually accessed by plotting the observed mean of log(antibody

level) for each infected and susceptible population and the overall mean as a function of age together with those figures obtained by fitting the SIR models (Figure 6). In both antibodies, the fitted means of infected and susceptible population are almost identical with the observed ones. However, for the fitted overall means, we can spot some deviations from the observed curves after age 30 for GLURP and 40 for CSP antibody.

3.2.2 SIRS



(a) GLURP **(b)** CSP



(c) GLURP **(d)** CSP

Figure 7: Univariate SIRS model: Upper panel: seroprevalence estimate; Lower panel: Goodness of fit. Black solid lines: mean of two subpopulations estimated using mixdist function. Red dashed lines: estimate mean of two subpopulations with frailty model. Black curve: observed overall mean using spline method. Red curve: fitted overall mean with frailty model

Likewise in SIR case, prevalence and force of infection are plotted against age for both antibodies in Figure 7. Compared to univariate SIR models, we observe the same pattern for the prevalences. The difference lies in the FOI estimates. While in GLURP, FOI of SIRS model is just slightly higher than that of SIR model, the difference is more pronounced in CSP case. For this antibody, from 40 years onward, FOI estimate surge nearly 100 times compared to the figure of SIR model. The goodness of fit is graphically displayed in Figure 7.

Parameter estimates of both models are presented in Table 2. While the majority of parameter estimates in both models are almost identical, we find a remarkable difference in variance of frailty term σ_f^2 estimates. For GLURP, σ_f^2 decreases from 1.35 in Model 1 to 0.87 in Model 2. For CSP, the estimate drops from 2.07 to 0.01 when we included replenishment rate in SIR model. In general, Model 2 yields slightly smaller AIC values compared to Model 1. Tests for heterogeneity based on 50:50 mixture of $\chi^2(0)$ and $\chi^2(1)$ yield p-value equal to 0.002 and 0.03 for CSP and GLURP respectively.

Table 2: Parameter estimates for SIR and SIRS univariate frailty gamma model

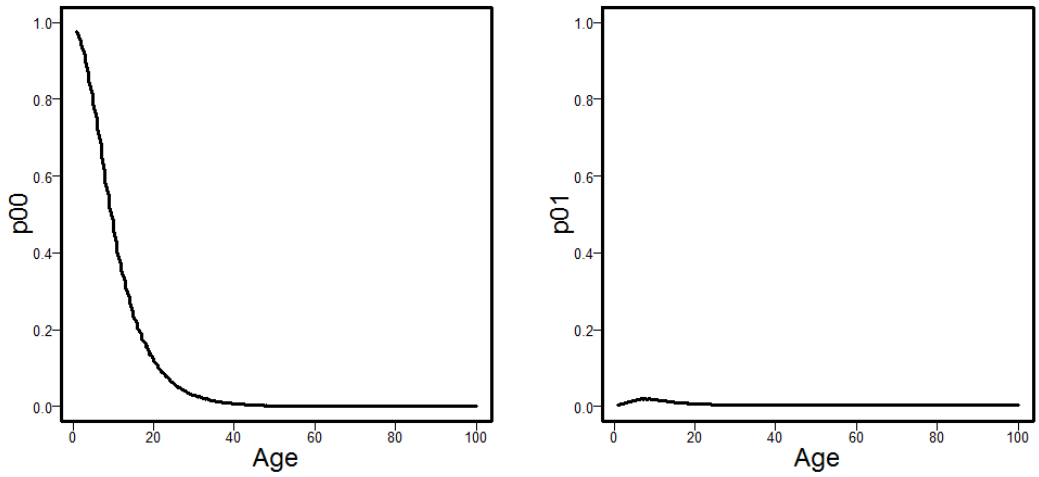
		Model 1				Model 2			
ID	Parameters	Estimates	lower	upper	AIC	Estimates	lower	upper	AIC
GLURP	σ_f^2	1.351	0.717	2.545	12585.9	0.866	0.457	1.640	12582.62
	σ					0.018	0.007	0.044	
	μ_0	3.684	3.658	3.711		3.690	3.663	3.717	
	δ_0	3.206	3.157	3.255		3.235	3.185	3.286	
	σ_0	0.491	0.469	0.514		0.496	0.474	0.519	
	σ_1	1.473	1.439	1.509		1.444	1.406	1.483	
	α	0.023	0.017	0.031		0.025	0.019	0.034	
	β	0.264	0.182	0.381		0.229	0.157	0.332	
CSP	σ_f^2	2.065	1.553	2.746	10969.3	0.011	8.5e-03	1.5e-02	10963.1
	σ					0.039	2.8e-08	5.5e+04	
	μ_0	3.254	3.225	3.283		3.255	3.23	3.28	
	δ_0	2.976	2.930	3.023		2.989	2.94	3.03	
	σ_0	0.655	0.633	0.679		0.657	0.634	0.680	
	σ_1	1.122	1.085	1.159		1.105	1.07	1.14	
	α	0.003	0.002	0.004		0.006	5.0e-03	7.7e-03	
	β	0.298	0.254	0.349		0.191	0.173	0.211	

Based on 95% confidence interval, all parameters are significantly different from 0. Nevertheless, the variability for recurrent rate σ of CSP is quite large and the lower bound almost equals zero. Therefore, when combining the two antibodies in a bivariate model, it is reasonable to proceed with SIRS model for GLURP and SIR model for CSP.

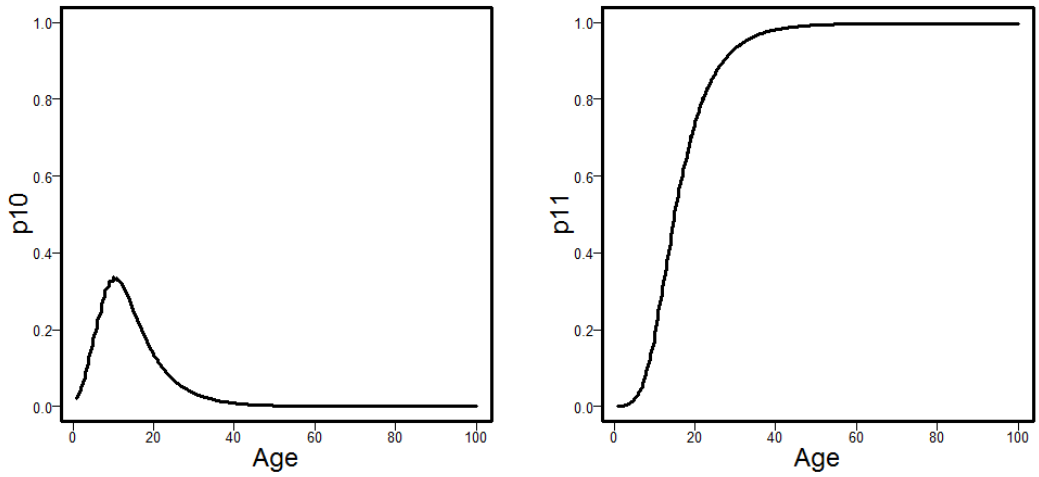
3.3 Bivariate

We now including the treatment effect in bivariate shared frailty model, assuming SIRS dynamic model for GLURP and SIR dynamic model for CSP. In Figure 8, estimated prevalences of individuals having one of the four possible serological profiles are presented. As can be seen, the prevalence of individuals who were seronegative with both antigens is decreasing with age. In contrast, we observe an increasing pattern in the prevalence of seropositives with respect to both antigens. The proportion of individuals

who is seropositive for GLURP and seronegative for CSP is higher than that of the reversed case.



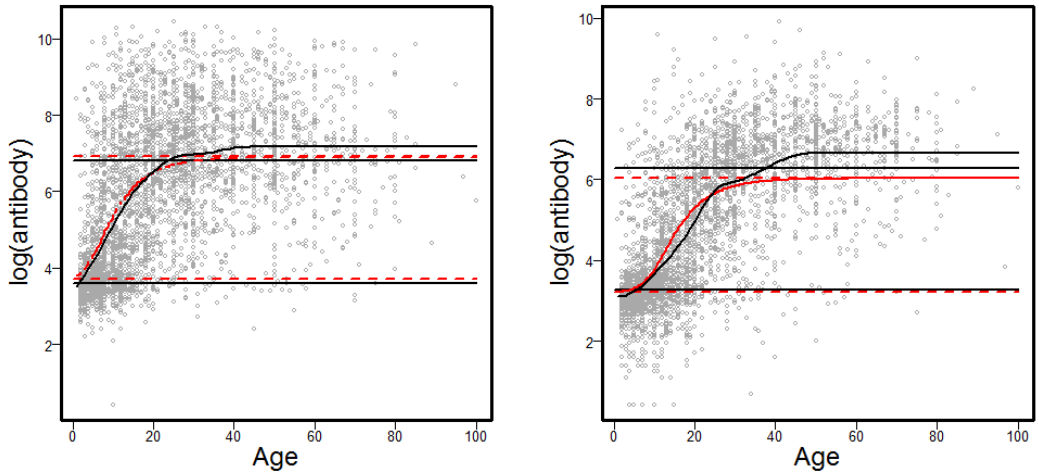
(a) p00 **(b)** p01



(c) p10 **(d)** p11

Figure 8: Bivariate shared SIRS-SIR model: seroprevalence estimate

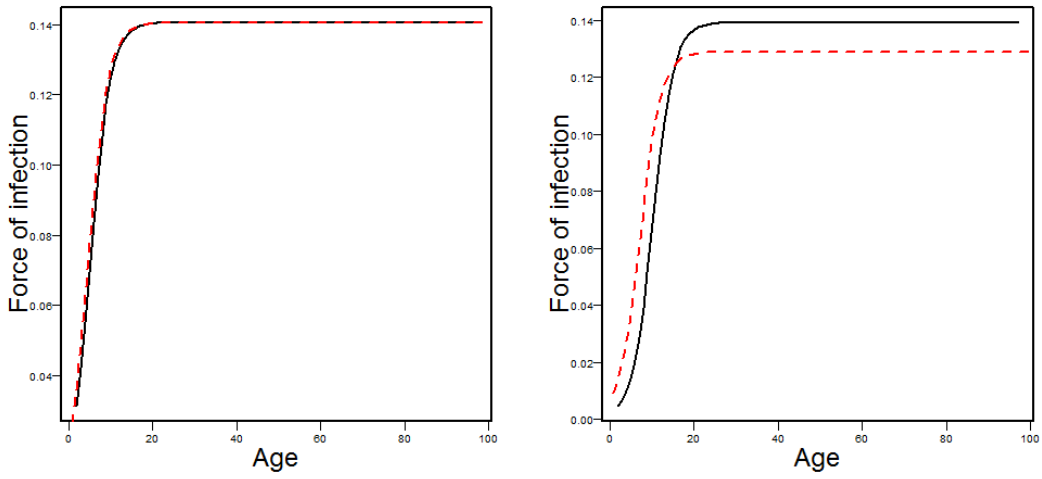
The marginal prevalence for each antibody are plotted against the observed prevalence to access goodness of fit in Figure 9. While with GLURP, a similar fit was obtained compared to univariate models, one can easily detect a larger deviation from the observed prevalence curve in CSP case.



(a) GLURP **(b)** CSP

Figure 9: Bivariate shared gamma frailty model: Goodness of fit. Black solid lines: mean of two subpopulations estimated using mixdist function. Red dashed lines: estimate mean of two subpopulations with frailty model. Black curve: observed overall mean using spline method. Red curve: fitted overall mean with frailty model

The obtained unconditional force of infection for each treatment arms are shown in Figure 10. In



(a) GLURP **(b)** CSP

Figure 10: Bivariate shared gamma frailty model: marginal Force of infection. Black solid line: FOI for control group. Red dashed line: FOI for treatment group

general, compared to univariate case, model with shared frailty term gives the same FOI pattern for both antibodies. However, while we obtain similar FOI estimate for CSP in comparison with its univariate counterpart model, the figure for GLURP in bivariate model is smaller than that of its SIRS univariate

model.

For most of the time, the FOI of treatment group in GLURP antibody is slightly higher than that of control group, whereas FOI of treatment group in CSP case is lower than that of control group. However, the differences are not significant since 95% interval of both treatment parameters contain 0. Parameter estimates of bivariate shared models are shown in Table 3. The shared frailty's variance σ_f^2 is 2.9, higher than frailty's variance of each antibody in univariate case. Though the interpretation of the frailty's variance in both cases are different. Based on AIC values, shared bivariate model outperforms univariate model, and shows an improvement in the overall goodness of fit. Indeed, AIC for bivariate model is 22668.09, whereas this figure for the two independent univariate models in combination is 23545.72.

Similar to shared frailty model, we fitted a correlated frailty model with SIRS dynamic for GLURP and SIR for CSP. When optimizing the model's likelihood, we did not obtain any improvement in the deviance's value compared to shared model. Furthermore, the algorithm fails in identifying a global maximization as the obtained inverse Hessian matrix keeps giving negative value on the diagonal for several different starting values.

Table 3: Bivariate shared gamma frailty model with treatment effect: Parameter estimates

Parameter	Estimate	Lower	Upper	AIC
				22668.09
GLURP				
σ_1	0.003	0.000	0.019	
μ_{01}	3.716	3.687	3.744	
δ_1	3.195	3.146	3.245	
α_1	0.016	0.012	0.021	
β_1	0.410	0.354	0.476	
β_1'	0.168	-0.085	0.420	
CSP				
μ_{02}	3.214	3.187	3.242	
δ_2	2.815	2.761	2.870	
α_2	0.002	0.001	0.003	
β_2	0.406	0.361	0.456	
β_2'	0.101	-0.164	0.366	
Common				
σ_f^2	2.916	2.448	3.474	
σ_{01}^2	0.252	0.229	0.277	
σ_{11}^2	2.070	1.970	2.176	
σ_{02}^2	0.360	0.331	0.391	
σ_{12}^2	1.511	1.402	1.629	
σ_{10}	0.239	0.177	0.300	
σ_{01}	0.357	0.197	0.516	
σ_{00}	0.087	0.066	0.107	
σ_{11}	0.711	0.630	0.793	

4 Discussion

Serological data is an important epidemiological source of information and has long been used in modeling the malaria transmission intensity. With the recent development of multiplex technique, multivariate serological data can be exploited to model the prevalence of the disease as well as incorporate heterogeneity of individuals through frailty models. The traditional frailty models are extended to account for disease that do not confer lifelong immunity by encompassing an extra replenishment rate. The methodology was illustrated on two antibodies: GLURP and CSP.

Our findings for univariate frailty model shown that models that account for recurrent infection slightly improved the goodness of fit for both antibodies. Notice that for CSP, the lower bound of 95% confidence interval for the replenishment rate σ is relatively close to zero, indicating that σ is on the borderline of statistical significance. Yet, the loglikelihood test for the heterogeneity using a mixture of $\chi^2(0)$ and $\chi^2(1)$ were significant in both antibodies. The unusual high FOI estimate in SIRS model for CSP might be the indicator of some boosting effects through exposure to replenishment individuals which occurred at 30 years onward in the population. Anther possible reason is that SIRS model is over-parameterized for CSP and the data at hand does not have enough evidence to estimate the replenishment rate which represented in a very wide confidence interval for σ.

Apart from the variance of frailty term, we found consistency in parameter estimates for the mean of each sero-status population as well as baseline FOI's parameters between two univariate frailty models. The SIRS model yields smaller estimate for the frailty variance compared to the SIR model where we assumed life-long immunity infection. This finding is no surprise as introducing one more parameter in the model will reduce the amount of variability that the frailty term has to capture.

While in univariate model, the frailty term expresses the heterogeneity among individual in the acquisition of disease, it has different meaning for bivariate models. Frailty term now imposes a correlation structure among infections or the dependency in acquisition of both antibodies, hence the frailty estimates are incomparable between the two models.

Treatment effect was incorporated in the model using proportional hazards assumption. For both antibodies, we found no evidence of difference in the two treatment arms. The obtained result matched our prediction since the data was taken from the first phase of the study when repellent had not clearly shown their effects. In shared frailty model, the values estimates of FOI for both antibodies are different with those obtained in each antibodies' corresponding dynamic univariate models. Together with the improvement in goodness of fit, combining antibodies in a bivariate model results in a better FOI estimate.

As an extension of shared frailty model, correlated model allows us to encompass a more flexible structure of correlation between frailty terms. However, the fact that no global maximization was obtained is an indication of model over-parameterization. Correlated frailty model can be fitted on another combination of antibodies to check for its applicability in modeling heterogeneity using a more complex correlation structure.

Although we have used Gompertz distribution to model the baseline FOI, others distribution can be applied as well, such as Lognormal, Weibull and Gamma... Moreover, we only fitted models with a constant replenishment rate σ, a more extended model that allow for an age-depend rate (eg. dichotomous

replenishment rate based on some cut-off point of time) should be advocated to obtain a more accurate estimate of the disease prevalence. Other alternatives could be to incorporate age and time dependency in the same model to relax the time homogeneous assumption or using different distribution (i.e. skewed log-normal) for each sero status population apart from the log-normal distribution. Furthermore, Abrams et al. (2014) also suggested some possibilities in further research with age-dependent frailty variables and more flexible correlation function between them.

In the context of the thesis, we adopted SIR and SIRS dynamic compartment to model the prevalence of seropostives. As already mentioned, in SIRS model, an individual is transferred from infected state to recovery state at a rate γ. In solving equation 1, we have assumed that the duration of infectious time is short which results in a small and neglectable infected proportion (I(a)). As a consequence, γ was not incorporated in the survival function $S(a)$ and hence we did not estimate this parameter. The assumption that we have made is quite strong with regard to malaria context since it would take years for patients to recover without treatment. Therefore, the assumption should be relaxed in order to develop a model that could estimate the recovery rate γ. On the other hand, in contrast with SIS model that usually used in many malaria studies, (SIR)SIRS model allows Recover state where individual confer a (lifelong) temporary immunity after getting infected. This state might be not appropriate in malaria disease as individual's antibody level can be boosted and individuals can be reinfected again. Therefore, SIS model can be used as well in modeling individual heterogeneity using frailty variable. With SIS model, we can get rid of the assumption about short infectious time, and can also derive the force of infection directly from the prevalence.

In order to validate the models, a simulation can be done to test for the applicability of the models in modeling the disease transmission.

5 Conclusion

In conclusion, the method that applied in this thesis allows us to estimate prevalence and force of infection base on serological data under several dynamic scenario, taken into account the individual heterogeneity in both univariate and bivariate fashions. The developed model can be adopted to test hypothesis of some covariate effects (eg. treatment). We conclude that the combination of antibodies in a bivariate model results in a better disease transmission parameter estimate. Further refinement and extension could be done in order to complete the new statistical tool to evaluate the disease transmission in pre-elimination areas. Finally, we need to emphasize that carrying parasite is different from carrying antibodies, and a decrease in disease prevalence does not necessary imply a declining in probability of being infected.

References

Abrams, Steven and Niel Hens (2014). “Modeling individual heterogeneity in the acquisition of recurrent infections: an application to parvovirus B19”. In: *Biostatistics*.

Ambrosino, Elena et al. (2010). “A multiplex assay for the simultaneous detection of antibodies against 15 Plasmodium falciparum and Anopheles gambiae saliva antigens”. English. In: *Malaria Journal* 9.1, pp. 1–12.

Centre for disease control and prevention (2014). *Malaria*. Accessed: 2014-03-23.

Cook J. Heidi Reid, Jennifer Iavro (2010). “Using serological measures to monitor changes in malaria transmission in Vanuatu”. In: *Malaria Journal* 9, p. 169.

Corran P. Coleman P., Riley E.; Drakeley C. (2004). “Serology: a robust indicator of malaria transmission intensity?” In: *Trends in parasitology* 23, pp. 575 –582.

Coutinho, F.A.B., E. Massad, L.F. Lopez, M.N. Burattini, C.J. Struchiner, and R.S. Azevedo-Neto (1999). “Modelling heterogeneities in individual frailties in epidemic models”. In: *Mathematical and Computer Modelling* 30.12, pp. 97 –115.

Cox, D.R. and D. Oakes (1996). *Analysis of Survival Data*. Monographs on statistics and applied probability. Chapman & Hall.

Elshal, Mohamed F. and J. Philip McCoy (2006). “Multiplex bead array assays: Performance evaluation and comparison of sensitivity to {ELISA}”. In: *Methods* 38.4. Methods for Analyzing Cytokines, pp. 317 –323.

Farrington, C. P., M. N. Kanaan, and N. J. Gay (2001). “Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 50.3, pp. 251–292.

Fouda, Genevieve G., Rose F. G. Leke, Carole Long, Pierre Druilhe, Ainong Zhou, Diane Wallace Taylor, and Armead H. Johnson (2006). “Multiplex Assay for Simultaneous Measurement of Antibodies to Multiple Plasmodium falciparum Antigens”. In: *Clinical and Vaccine Immunology* 13.12, pp. 1307–1313.

Gay, NIGEL J. (1996). “Analysis of serological surveys using mixture models: application to a survey of Parvovirus B19”. In: *Statistics in Medicine* 15.14, pp. 1567–1573.

Hens Shkedy Z., Aerts M.and Faes C., Van Damme P., and Beutels P. (2012). *Modeling Infectious Disease Parameters Based on Serological and Social Contact Data*. London: Springer.

Hens, N., A. Wienke, M. Aerts, and G. Molenberghs (2009). “The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data”. In: *Statistics in Medicine* 28.22, pp. 2785–2800.

Kondrashin, Anatoly, Alla Baranova, Elizabeth Ashley, Judith Recht, Nicholas White, and Vladimir Sergiev (2014). “Mass primaquine treatment to eliminate vivax malaria: lessons from the past”. In: *Malaria Journal* 13.1, p. 51.

Marcus, Bernard (2009). *Deadly Diseases and Epidemics Series- Malaria*. New York: Chelsea House.

Morley, John E., H. Mitchell Perry, and Douglas K. Miller (2002). “Editorial: Something About Frailty”. In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 57.11, pp. M698–M704.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Vaupel, JamesW., KennethG. Manton, and Eric Stallard (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality". English. In: *Demography* 16.3, pp. 439–454.

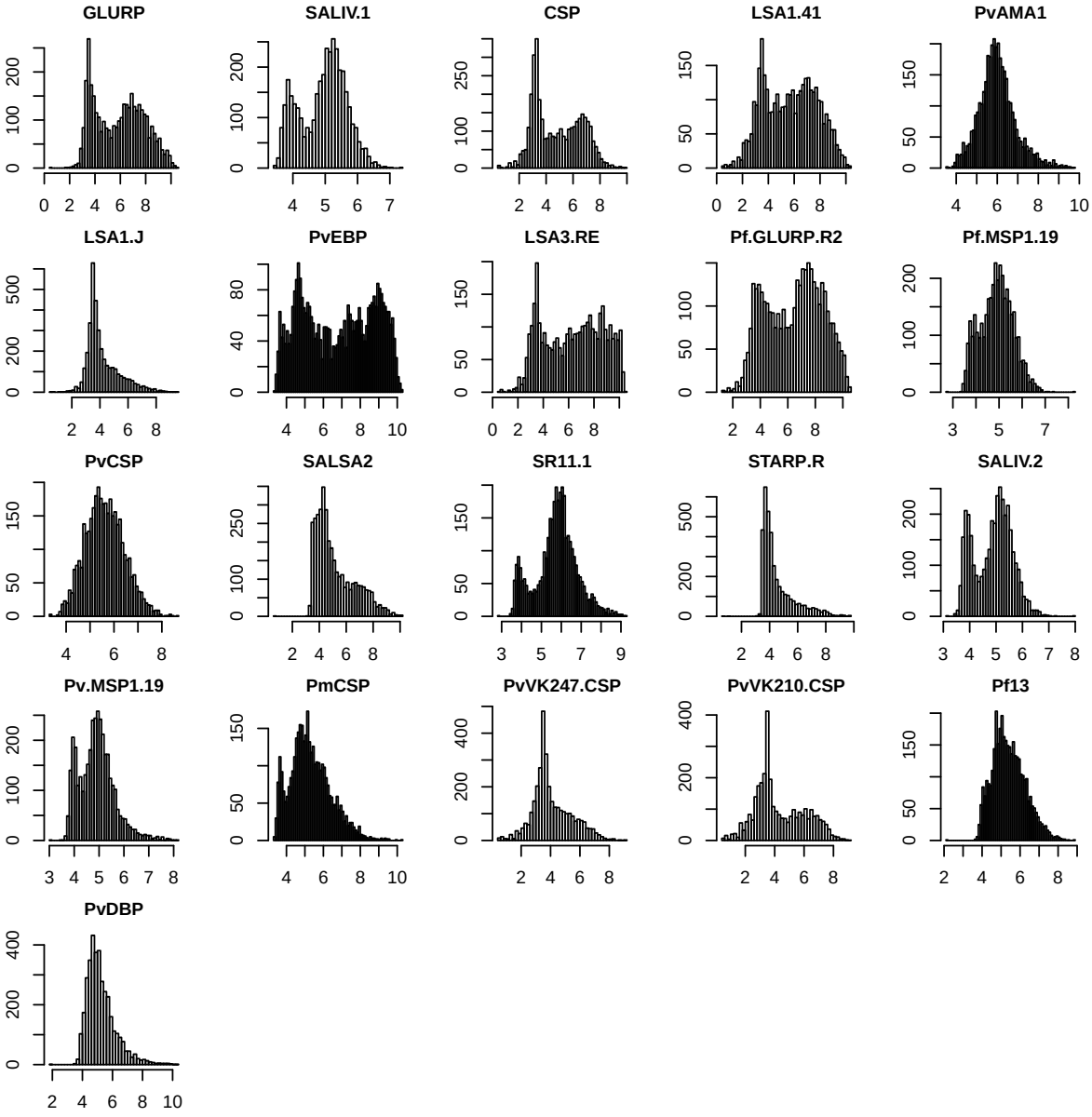
Webster, H. K., J. B. Gingrich, C. Wongsrichanalai, S. Tulyayon, A. Suvarnamani, P. Sookto, and B. Permapanich (1992). "Circumsporozoite Antibody as a Serologic Marker of Plasmodium Falciparum Transmission". In: *The American Journal of Tropical Medicine and Hygiene* 47.4, pp. 489–497.

Wienke, Andreas (2010). *Frailty Models in Survival Analysis*. Chapman and Hall/CRC.

World Health Organization (2013). *World Malaria Report 2013*.

A Appendix

EDA of 21 antibodies



B R code

SIR model

```
th.SIR.Gomp<-function(a,al1,startpar){
  ptm<-proc.time()
  qprocSIR<-function(a=a,al1=al1,par){
    logsigma2.1f = par[1]
    logmu0 = par[2]
    logdelta1 = par[3]
    logsigma0 = par[4]
    logsigma1 = par[5]
    alphaeta = par[6]
    betaeta = par[7]

    mu0 = exp(logmu0)
    mu1 = mu0+exp(logdelta1)
    sigma0 = exp(logsigma0)
    sigma1 = exp(logsigma1)
    sigma2.1f = exp(logsigma2.1f)
    if (sigma2.1f > 10000) {sigma2.1f<-10000}
    if (sigma2.1f == 0) {sigma2.1f<-1e-16}

    # Parametric function for FOI:Gompertz FOI=alpha*exp(beta*age) (alpha, beta>0)
    alpha=exp(alphaeta)
    beta=exp(betaeta)
    cfoi= function(x){
      alpha/beta*(exp(x*beta)-1)
    }

    # Calculate S(a)
    laplace1<-function(s) {return((1+(sigma2.1f*s))**(-1/sigma2.1f))}
    prev1<-rep(NA,length(a))
    ll2<-rep(NA,length(a))
    save=NULL
    agrid<-a1
    for (i in 1:length(agrid)){
      S<-laplace1(cfoi(agrid[i]))
      save[i]=S
      prev1[i]<-(1-S)
      if (prev1[i]<0) {prev1[i]<-0}
      if (prev1[i]>1) {prev1[i]<-1}
      ####
      ll2[i]<-log((1-prev1[i])*dnorm(log(al1[i]+30),mean=mu0,sd=sigma0)+prev1[i]*dnorm(log(
    al1[i]+30),mean=mu1,sd=sigma1)+1e-8)
      ####
    }# end of for

    return(list(save=save, prev1=prev1,
                ll=-2*(sum(ll2))))
  }

  # take the -2loglikelihood
  qproc.fitter<-function(par){
    qproc.ll<-qprocSIR(a=a,al1=al1,par)$ll
    return(qproc.ll)}
  startpar<-c(startpar)
  # Optimized the log likelihood with some initial values
```

```
  q.result<-nlm(qproc.fitter,startpar,hessian=T,iterlim=200,print.level=2)
  #q.result<-optim(startpar2, qproc.fitter,hessian=T)

  # Get the result and run qproc to get others par
  result.global<-qprocSIR(a=a,al1=al1,q.result$estimate)
  runtime<-((proc.time()-ptm)/60)[1]
  return(list(sigma2.1f=exp(q.result$estimate[2]),
              sigmahat=exp(q.result$estimate[1]),logmu0=q.result$estimate[3],logdelta1=
    q.result$estimate[4],
              logsigma0=q.result$estimate[5],logsigma1=q.result$estimate[6],alphaeta=
    q.result$estimate[7],
              betaeta=q.result$estimate[8],
              hess=q.result$hessian,deviance=q.result$minimum,aic=q.result$minimum+(2*
    length(q.result$estimate)),
              bic=q.result$minimum+(log(length(al1))*3),
              foi10=result.global$foi10,prev1=result.global$prev1,
              convergence=q.result$code,runtime=runtime))

}
```

SIRS model

```
## SIRS with Gompz
th.SIRS.Gomp<-function(a,al1,startpar){
  ptm<-proc.time()
  qprocSIRS<-function(a=a,al1=al1,par){
    logsigma = par[1]
    logsigma2.1f = par[2]
    logmu0 = par[3]
    logdelta1 = par[4]
    logsigma0 = par[5]
    logsigma1 = par[6]
    alphaeta = par[7]
    betaeta = par[8]

    sig = exp(logsigma)
    mu0 = exp(logmu0)
    mu1 = mu0+exp(logdelta1)
    sigma0 = exp(logsigma0)
    sigma1 = exp(logsigma1)
    sigma2.1f = exp(logsigma2.1f)
    if (sig > 10000) {sig<-10000}
    if (sigma2.1f > 10000) {sigma2.1f<-10000}
    if (sigma2.1f == 0) {sigma2.1f<-1e-16}

    # Parametric function for FOI:Gompertz FOI=alpha+exp(beta*age) (alpha, beta>0)
    alpha=exp(alphaeta)
    beta=exp(betaeta)
    cfoi= function(x){
      cfoii10<-alpha/beta*(exp(x*beta)-1)
    }

    # Calculate S(a)
    laplace1<-function(s) {return((1+(sigma2.1f*s))**(-1/sigma2.1f))}
    prev1<-rep(NA,length(a))
    ll2<-rep(NA,length(a))
    agrid<-a
    for (i in 1:length(agrid)){
```

```
      term1<-exp(-sig*agrid[i])
      term2<-laplace1(cfoi(agrid[i]))

      term31<-matrix(sig*exp(-(sig*agrid[i])+(sig*cumsum(rep(1,agrid[i])))), nrow = 1)
      term32<-matrix(laplace1(cfoi(agrid[i])- cfoi(1:agrid[i])),ncol = 1)
      term3<-term31%*%term32

      S<-(term1*term2) + term3
      prev1[i]<-(1-S)
      if (prev1[i]<0) {prev1[i]<-0}
      if (prev1[i]>1) {prev1[i]<-1}
      ####
      ll2[i]<-log((1-prev1[i])*dnorm(log(al1[i]+30),mean=mu0,sd=sigma0)+prev1[i]*dnorm(log(
    al1[i]+30),mean=mu1,sd=sigma1)+1e-8)
      ####
    }# end of for

    return(list(prev1=prev1,
                ll=-2*(sum(ll2))))
  }
  #qproc(a1,z1,startpar2)$ll
  # take the -2loglikelihood
  qproc.fitter<-function(par){
    qproc.ll<-qprocSIRS(a=a,al1=al1,par)$ll
    return(qproc.ll)}
  startpar<-c(startpar)
  # Optimized the log likelihood with some initial values
  q.result<-nlm(qproc.fitter,startpar,hessian=T,iterlim=400,print.level=2)
  #q.result<-optim(startpar2, qproc.fitter,hessian=T)

  # Get the result and run qproc to get others par
  result.global<-qprocSIRS(a=a,al1=al1,q.result$estimate)
  runtime<-((proc.time()-ptm)/60)[1]
  return(list(sigma2.1f=exp(q.result$estimate[2]),
              sigmahat=exp(q.result$estimate[1]),logmu0=q.result$estimate[3],logdelta1=
    q.result$estimate[4],
              logsigma0=q.result$estimate[5],logsigma1=q.result$estimate[6],alphaeta=
    q.result$estimate[7],
              betaeta=q.result$estimate[8],
              hess=q.result$hessian,deviance=q.result$minimum,aic=q.result$minimum+(2*
    length(q.result$estimate)),
              bic=q.result$minimum+(log(length(al1))*3),
              foi10=result.global$foi10,prev1=result.global$prev1,
              convergence=q.result$code,runtime=runtime))

}
```

Shared bivariate SIRS-SIR model

```
qproc.share.treat.sirs.sir<- function(a,z1,z2,treat,par){

  #
  logsigma1 = par[1]    # replemishment rate 1
  logsigma2.f = par[2] # for laplace

  logmu01 = par[3] # mu of S1
  logmu02 = par[4] # mu of S2
  logdelta1 = par[5]
```

```
logdelta2 = par[6]

logsigma2.01 = par[7] # sigma^2 of S1
logsigma2.11 = par[8] # sigma^2 of I1
logsigma2.02 = par[9] # sigma^2 of S2
logsigma2.12 = par[10] # sigma^2 of I2

sigma10 = par[11] # sigma of I1S2
sigma01 = par[12] # sigma of S1I2
sigma00 = par[13] # sigma of S1S2
sigma11 = par[14] # sigma of I1I2

alphaeta1 = par[15]
alphaeta2 = par[16]
betaeta1 = par[17]
betaeta2 = par[18]
betatrm1 =  par[19]
betatrm2 =  par[20]
sig1 = exp(logsigma1)
sigma2.f = exp(logsigma2.f)

mu01 = exp(logmu01)
mu02 = exp(logmu02)
mu11 = mu01+exp(logdelta1)
mu12 = mu02+exp(logdelta2)

sigma2.01 = exp(logsigma2.01)
sigma2.11 = exp(logsigma2.11)
sigma2.02 = exp(logsigma2.02)
sigma2.12 = exp(logsigma2.12)


Sigma10 = matrix(c(sigma2.11,sigma10,sigma10,sigma2.02),2,2, byrow = T)
Sigma01 = matrix(c(sigma2.01,sigma01,sigma01,sigma2.12),2,2, byrow = T)
Sigma00 = matrix(c(sigma2.01,sigma00,sigma00,sigma2.02),2,2, byrow = T)
Sigma11 = matrix(c(sigma2.11,sigma11,sigma11,sigma2.12),2,2, byrow = T)

## FOI: Gomperzt
alpha1=exp(alphaeta1)
beta1=exp(betaeta1)
alpha2=exp(alphaeta2)
beta2=exp(betaeta2)
cfoi1= function(x){
  cfoii10<-alpha1/beta1*(exp(x*beta1)-1)
}
cfoi2= function(x){
  cfoii10<-alpha2/beta2*(exp(x*beta2)-1)
}
## calculate S
laplace<-function(s) {return((1+(sigma2.f*s))**(-1/sigma2.f))}

S1<-rep(NA,length(a))
S2<-rep(NA,length(a))
S12<-rep(NA,length(a))
p11<-rep(NA,length(a))
p10<-rep(NA,length(a))
p01<-rep(NA,length(a))
```

```
p00<-rep(NA,length(a))
ll<-rep(NA,length(a))
agrid<-a
## S1(SIRS) and S2(SIR)
for (i in 1:length(agrid)){
  term1<-exp(-sig1*agrid[i])
  term2<-laplace(exp(betatrm1*treat[i])*cfoi1(agrid[i]))

  term31<-matrix(sig1*exp(-(sig1*agrid[i])+(sig1*cumsum(rep(1,agrid[i])))), nrow = 1)
  term32<-matrix(laplace(exp(betatrm1*treat[i])*(cfoi1(agrid[i])- cfoi1(1:agrid[i]))),
  ncol = 1)
  term3<-term31%*%term32

  S1[i]<-(term1*term2) + term3
  S2[i]<-laplace(exp(betatrm2*treat[i])*cfoi2(agrid[i]))

  term1.S12<-exp(-sig1*agrid[i])
  term2.S12<-laplace(exp(betatrm1*treat[i])*cfoi1(agrid[i])+exp(betatrm2*treat[i])*cfoi2(
  agrid[i]))

  if (floor(agrid[i])!=0){
    term31.S12<-matrix(sig1*exp(-(sig1*agrid[i])+(sig1*cumsum(rep(1,floor(agrid[i]))))),
  nrow = 1)
    term32.S12<-matrix(laplace(exp(betatrm1*treat[i])*(cfoi1(agrid[i])-cfoi1(1:agrid[i]))
  +exp(betatrm2*treat[i])*cfoi2(agrid[i])),ncol = 1)
    term3.S12<-term31.S12%*%term32.S12}
  else {term3.S12<-0}

  #term4.S12<-(sig1*(agrid[i]-floor(agrid[i])))*laplace(c(0,cumsum(foii20))[floor(agrid[i
  ])+1]+(foii20[floor(agrid[i])+1]*(agrid[i]-floor(agrid[i]))))
  S12[i]<-(term1.S12*term2.S12) + term3.S12 #+ term4.S12

  p11[i]<-1-S1[i]-S2[i]+S12[i]
  p10[i]<-S2[i]-S12[i]
  p01[i]<-S1[i]-S12[i]
  p00[i]<-S12[i]

  if (p11[i]<0) {p11[i]<-0}
  if (p11[i]>1) {p11[i]<-1}
  if (p10[i]<0) {p10[i]<-0}
  if (p10[i]>1) {p10[i]<-1}
  if (p01[i]<0) {p01[i]<-0}
  if (p01[i]>1) {p01[i]<-1}
  if (p00[i]<0) {p00[i]<-0}
  if (p00[i]>1) {p00[i]<-1}


  ll[i]<-log(  p11[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu11,mu12),
  sigma = Sigma11)+
                p10[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu11,mu02),
  sigma = Sigma10)+
                p01[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu01,mu12),
  sigma = Sigma01)+
                p00[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu01,mu02),
  sigma = Sigma00)+1e-8)

}# end of for
```

```
  return(list(ll=-2*sum(ll), S1=S1,S2=S2,S12=S12,p11=p11,p10=p10,p01=p01,p00=p00))
}# end qproc

qproc.treat.fitter<-function(start){
  qproc.crit<-qproc.share.treat.sirs.sir(a,z1,z2,treat, start)$ll
  return(qproc.crit)}
q.result.treat_n<-nlm(qproc.treat.fitter,start,hessian=T,iterlim=1000,print.level = 2)
result.share.treat.sirs.sir<-qproc.share.treat.sirs.sir(a, z1, z2, treat, q.result.treat_n$
    estimate)
aic=q.result.treat_n$minimum+2*length(q.result.treat_n$estimate)
```

Correlated bivariate SIRS-SIR model

```
library(cubature)
library(mvtnorm)
qproc.cor<-function(a=a,z1=z1,z2=z2,par){
  logsigma1 = par[1]    # replemishment rate 1
  logk0 = par[2]
  logk1 = par[3]
  logk2 = par[4]
  k0<-exp(logk0)
  k1<-exp(logk1)
  k2<-exp(logk2)
  sigma2.1f<-1/(k0+k1)
  sigma2.2f<-1/(k0+k2)
  rho12<-k0/(sqrt((k0+k1)*(k0+k2)))
  sig1<- exp(logsigma1)
  #sig<-rep(sig1,100)

  logmu01 = par[5] # mu of S1
  logmu02 = par[6] # mu of S2
  logdelta1 = par[7]
  logdelta2 = par[8]

  logsigma2.01 = par[9] # sigma^2 of S1
  logsigma2.11 = par[10] # sigma^2 of I1
  logsigma2.02 = par[11] # sigma^2 of S2
  logsigma2.12 = par[12] # sigma^2 of I2

  sigma10 = par[13] # sigma of I1S2
  sigma01 = par[14] # sigma of S1I2
  sigma00 = par[15] # sigma of S1S2
  sigma11 = par[16] # sigma of I1I2

  alphaeta1 = par[17]
  alphaeta2 = par[18]
  betaeta1 = par[19]
  betaeta2 = par[20]

  mu01 = exp(logmu01)
  mu02 = exp(logmu02)
  mu11 = mu01+exp(logdelta1)
  mu12 = mu02+exp(logdelta2)

  sigma2.01 = exp(logsigma2.01)
  sigma2.11 = exp(logsigma2.11)
  sigma2.02 = exp(logsigma2.02)
```

```
sigma2.12 = exp(logsigma2.12)


Sigma10 = matrix(c(sigma2.11,sigma10,sigma10,sigma2.02),2,2, byrow = T)
Sigma01 = matrix(c(sigma2.01,sigma01,sigma01,sigma2.12),2,2, byrow = T)
Sigma00 = matrix(c(sigma2.01,sigma00,sigma00,sigma2.02),2,2, byrow = T)
Sigma11 = matrix(c(sigma2.11,sigma11,sigma11,sigma2.12),2,2, byrow = T)



## FOI: Gomperzt
alpha1 = exp(alphaeta1)
beta1 = exp(betaeta1)
alpha2 = exp(alphaeta2)
beta2 = exp(betaeta2)
cfoi1 = function(x){
  cfoii10<-alpha1/beta1*(exp(x*beta1)-1)
}
cfoi2= function(x){
  cfoii10<-alpha2/beta2*(exp(x*beta2)-1)
}

  laplace.Y0<-function(s) {return((1+s)**(-k0))}
  laplace.Y1<-function(s) {return((1+s)**(-k1))}
  laplace.Y2<-function(s) {return((1+s)**(-k2))}
  S1<-rep(NA,length(a))
  S2<-rep(NA,length(a))
  S12<-rep(NA,length(a))
  p11<-rep(NA,length(a))
  p10<-rep(NA,length(a))
  p01<-rep(NA,length(a))
  p00<-rep(NA,length(a))
  ll<-rep(NA,length(a))
  for (i in 1:length(a)){
    term1<-exp(-sig1*a[i])
    term2<-laplace.Y0(sigma2.1f*cfoi1(a[i]))*laplace.Y1(sigma2.1f*cfoi1(a[i]))

      term31<-matrix(sig1*exp(-(sig1*a[i])+(sig1*cumsum(rep(1,a[i])))), nrow = 1)
      term32<-matrix(laplace.Y0(sigma2.1f*(cfoi1(a[i])-cfoi1(1:a[i])))*
                       laplace.Y1(sigma2.1f*(cfoi1(a[i])-cfoi1(1:a[i]))),ncol=1)
      term3<-(term31%*%term32)

    S1[i]<-(term1*term2)+term3
    S2[i]<-laplace.Y0(sigma2.2f*cfoi2(a[i]))*laplace.Y2(sigma2.2f*cfoi2(a[i]))

    term1.S12<-exp(-sig1*a[i])
    term2.S12<-laplace.Y0((sigma2.1f*cfoi1(a[i]))+
                            (sigma2.2f*cfoi2(a[i])))
    term3.S12<-laplace.Y1((sigma2.1f*cfoi1(a[i])))
    term4.S12<-laplace.Y2((sigma2.2f*cfoi2(a[i])))

      term51.S12<- matrix(sig1*exp(-(sig1*a[i])+(sig1*cumsum(rep(1,a[i])))), nrow = 1)
      term53.S12<-matrix(laplace.Y0(sigma2.2f*cfoi2(a[i])+(sigma2.1f*(cfoi1(a[i])-cfoi1
  (1:a[i]))))*
                           laplace.Y2(sigma2.2f*cfoi2(a[[i]]))*
                           laplace.Y1(sigma2.1f*(cfoi1(a[i])-cfoi1(1:a[i]))),ncol=1)
      term5.S12<-term51.S12%*%term53.S12
```

```
      S12[i]<-(term1.S12*term2.S12*term3.S12*term4.S12)+term5.S12
      p11[i]<-1-S1[i]-S2[i]+S12[i]
      p10[i]<-S2[i]-S12[i]
      p01[i]<-S1[i]-S12[i]
      p00[i]<-S12[i]
      p11[i] = max(1e-8,p11[i]); p11[i] = min(1-1e-8,p11[i]);
      p10[i] = max(1e-8,p10[i]); p10[i] = min(1-1e-8,p10[i]);
      p01[i] = max(1e-8,p01[i]); p01[i] = min(1-1e-8,p01[i]);
      p00[i] = max(1e-8,p00[i]); p00[i] = min(1-1e-8,p00[i]);
      ll[i]<-log(  p11[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu11,mu12),
    sigma = Sigma11)+
                     p10[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu11,mu02)
    , sigma = Sigma10)+
                     p01[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu01,mu12)
    , sigma = Sigma01)+
                     p00[i]*dmvnorm(x = c(log(z1[i]+30),log(z2[i]+30)), mean = c(mu01,mu02)
    , sigma = Sigma00)+1e-8)

    }
    #print(-2*sum(ll))
    return(list(dev=-2*(sum(ll)),crit=-sum(ll),
                S1=S1,S2=S2,S12=S12,
                p11=p11,p10=p10,p01=p01,p00=p00,
                sigma2.1f=sigma2.1f,sigma2.2f=sigma2.2f,rho12=rho12))
  }

qproc.fitter.cor<-function(theta){
    qproc.crit<-qproc.cor(a, z1, z2, theta )$crit
    return(qproc.crit)}
qproc.fitter.cor(start.cor)
q.result.cor<-nlm(qproc.fitter.cor,start.cor,hessian=T,iterlim=1000, print.level = 2)
result.cor.global<-qproc.cor(a, z1, z2, q.result.cor$estimate)
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Statistical analysis of multiplex serological data: new tools to evaluate malaria transmission in pre-elimination settings.

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Le Thi Phuong, Thao

Datum: **10/09/2014**