

# Age and seasonal variation in the transition rates and detectability of *Plasmodium falciparum* malaria

W. SAMA<sup>1\*</sup>, S. OWUSU-AGYEI<sup>2</sup>, I. FELGER<sup>3</sup>, K. DIETZ<sup>4</sup> and T. SMITH<sup>1</sup>

<sup>1</sup> Department of Public Health and Epidemiology, and

<sup>3</sup> Department of Medical Parasitology and Infection Biology, Swiss Tropical Institute, Socinstrasse 57, Postfach CH-4002, Basel, Switzerland

<sup>2</sup> Kintampo Health Research Centre, Ghana Health Service, P.O. Box 200, Kintampo, Ghana

<sup>4</sup> Institut für Medizinische Biometrie, Westbahnhofstrasse 55, D-72070 Tübingen, Germany

(Received 2 May 2005; revised 20 June 2005; accepted 20 June 2005; first published online 13 September 2005)

## SUMMARY

The effect of acquired immunity on the duration of *Plasmodium falciparum* infections is unclear, although this is an important term in models of malaria transmission. It is problematical to determine the duration of infections because of the difficulty of distinguishing persisting infections from new ones, and because parasite densities are often transiently below the limit of detection. We recently developed a dynamic model for infection incidence, clearance and detection of multiple genotype *P. falciparum* infections and fitted it to a panel dataset from a longitudinal study in Northern Ghana. We now extend this model to allow for seasonal and age variation in infection rates and also age dependence in clearance and in detectability of infections. These models indicate that there is seasonal variation in the infection rate, and age dependence in detectability. The best fitting models had no age dependence in infection or clearance rates, suggesting that acquired immunity mainly affects detectability.

Key words: *Plasmodium falciparum*, infection rate, clearance rate, detectability, seasonal variation, acquired immunity, model.

## INTRODUCTION

The dynamics of *Plasmodium falciparum* malaria infections in endemic areas are crucial determinants of the effects of preventative interventions but are difficult to study because of the challenge of distinguishing persisting infections from new ones, and because parasite densities are often transiently below the limit of detection.

Where infectious agents are endemic, levels of immunity generally increase with age, and it is therefore to be expected that the duration of infections will be lower in older hosts than in younger ones. While this pattern has been assumed in many models of malaria transmission (Dietz, Molineaux and Thomas, 1974; Aron, 1988), the empirical evidence for it is weak (Walton, 1947; Kitua *et al.* 1996; Smith *et al.* 1999*a, b*; Smith and Vounatsou, 2003; Sama, Killeen and Smith, 2004).

We recently reported a model (Sama *et al.* 2005) to estimate infection and recovery rates (and hence the duration of infections) from repeated observations of the presence or absence of *P. falciparum* genotypes in the same group of individuals in the

Kassena-Nankana district (KND) in Northern Ghana. In developing this approach we treated the infection and recovery rates as constant. It was assumed that the laboratory test used to detect the infectious agent has imperfect detectability, but the detectability was assumed to be constant across the whole population. We also assumed that these parameters are the same for all genotypes.

We now extend our previous model (Sama *et al.* 2005) by explicitly parameterizing the detectability as well as the infection and recovery processes as functions of age. We also extend our model to allow for seasonal variation in infection rates.

## MATERIALS AND METHODS

### Field surveys

The KND, is a highly endemic malarious area showing a peak in transmission during the short wet season between May and September and corresponding seasonality in prevalence and clinical incidence (Binka *et al.* 1994; Baird *et al.* 2002; Koram *et al.* 2003). We analysed *P. falciparum* genotype data of 69 individuals (Sama *et al.* 2005), ranging in age from 1 month to 84 years. Each individual contributed 6 blood samples, with the first survey carried out in July 2000 and the remaining surveys at regular intervals of 2 months subsequently, each survey lasting for 8 or 9 days. Samples were analysed for the

\* Corresponding author: Department of Public Health and Epidemiology, Swiss Tropical Institute, Socinstrasse 57, Postfach CH-4002, Basel, Switzerland. Tel: +41 (0) 61 284 82 82. Fax: +41 (0) 61 284 81 05. E-mail: Wilson.Sama@unibas.ch

presence or absence of *P. falciparum* merozoite surface protein 2 genotypes using a PCR-RFLP (Polymerase Chain Reaction-Restriction Fragment Length Polymorphism) technique (Felger *et al.* 1999). There were a total of 70 *mSP2* genotypes observed, 33 belonging to the FC27 and 37 to the 3D7 allelic family.

As an example of the structure of the dataset analysed, consider an individual observed with 4 genotypes during the entire 6 surveys, whose data was represented by the following 4 sequences: (001001), (010000), (001110), (111100), corresponding to the 4 genotypes observed and indicating that the first genotype was observed during the third and sixth survey, while the second genotype was observed only during the second survey, and so on. The complete dataset consisted of 827 observed sequences of this nature.

*Model of parasite dynamics*

We previously used the immigration-death model corresponding to that originally proposed by Macdonald (1950) to describe  $n(a)$ , the expected number of distinct genotypes (or the expected true multiplicity of infections) within an individual of age  $a$  by the following equation.

$$\frac{dn(a)}{da} = \lambda - \mu n(a) \tag{1}$$

where  $\lambda$  is the infection rate, that is the rate at which new infections are acquired, and  $\mu$  is the clearance rate, that is the rate at which infections are cleared, both assumed to be homogeneous across the population. We now allow these parameters  $\lambda, \mu$ , to vary by age and season,  $t$ , so that equation 1 takes the form.

$$\frac{dn(a, t)}{da} = \lambda(a, t) - \mu(a)n(a, t) \tag{2}$$

The general solution of this equation is given by:

$$n(a, t) = e^{-\int \mu(a) da} \left[ K + \int \lambda(a, t) e^{\int \mu(a) da} da \right] \tag{3}$$

where  $K$  is the constant of integration obtained by substituting the initial condition  $n(0, t) = 0$ .

(a) *Infection process.* We evaluate 3 alternative forms for  $\lambda(a, t)$

- (i)  $\lambda(a, t) = \lambda$ , i.e. we treat the infection rate as constant.
- (ii)  $\lambda(a, t) = \beta_0 e^{\beta_1 a}$ , i.e. we assume that the infection rate is a monotonic function of age, where  $\beta_0$  is the infection rate at birth while  $\beta_1$  is the change in infection rate (on the logarithmic scale) for a unit increase in age.
- (iii)  $\lambda(a, t) = \lambda_i$ , i.e. seasonal variation in infection rates.  $i = 1, 2, \dots, 6$  indexes the two month

period of the year at time  $t$ , corresponding to the inter-survey interval and the parameters  $\lambda_1, \lambda_2, \dots, \lambda_6$  account for seasonal variation in the infection rate, with the rates treated as constant within each inter-survey period. In this model  $\lambda(a, t)$  is assumed to follow a recurring annual cycle and the same vector of parameters,  $\lambda_i$ , is applied for each year of life of the individual preceding the observation period.

(b) *Clearance process.* We evaluate 2 alternative forms for  $\mu(a)$

- (i)  $\mu(a) = \mu$
- (ii)  $\mu(a) = \mu_0 e^{\mu_1 a}$

$\mu_0$  is the clearance rate at birth while  $\mu_1$  is the change in clearance rate (on the logarithmic scale) for a unit increase in age.

(c) *Observation process.* The observed sequences may differ from true sequences because the observation process is imperfect. We assume one of the following forms for the detectability,  $s$ , the probability of detecting an infection in a blood sample conditional on it being present in the host.

- (i) Constant detectability,  $s(a) = s$ ;
- (ii) Age-dependent detectability,  $\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$ , where  $\bar{a}$  is the mean age (20.12) and  $s_0$  and  $s_1$  are parameters to be estimated.

The prediction for the observed mean multiplicity (or simply the expected multiplicity) is obtained by multiplying equation 3 by  $s(a)$ , that is:

$$\bar{n}(a, t) = s(a) e^{-\int \mu(a) da} \left[ K + \int \lambda(a, t) e^{\int \mu(a) da} da \right] \tag{4}$$

The expected (net) true number of infections acquired during the interval  $t$  to  $t + \tau$  in an individual of initial age  $a$  is then  $m(a + \tau, t + \tau)$  where:

$$\frac{dm(a, t)}{da} = \lambda(a, t) - \mu(a)m(a, t) \tag{5}$$

and the initial condition is  $m(a, t) = 0$  (Sama *et al.* 2005). In this paper, all units of measurement for time and age are years and for rates are year<sup>-1</sup>.

We assumed that re-infection with a genetically identical parasite clone is a rare event, and thus that observed sequences containing negative samples between 2 positive samples (for example as in the third sample of the sequence 010100 and as in the third and fourth samples of the sequence 110010) invariably result from failure to detect the infection. We also assumed that our test had perfect specificity. By maintaining these two assumptions, the derivation of the Poisson-likelihood for the frequency of each observed sequence is the same as previously described (Sama *et al.* 2005). The solutions of equations 2 and 5 are important components in this

Table 1. Different models evaluated

Model	Clearance rate, $\mu(a)$	Infection rate, $\lambda(a, t)$	Detectability, $s(a)$	AIC*	p†
M1	$\mu$	$\lambda$	$s(a) = s$	2709.6	3
M2	$\mu$	$\lambda$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2599.3	4
M3	$\mu$	$\lambda_{i_i}, i = 1, 2, \dots, 6$	$s(a) = s$	2644.3	8
M4	$\mu_0 e^{\mu_1 a}$	$\lambda$	$s(a) = s$	2620.5	4
M5	$\mu$	$\beta_0 e^{\beta_1 a}$	$s(a) = s$	2643.9	4
M6	$\mu$	$\lambda_{i_i}, i = 1, 2, \dots, 6$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2528.3	9
M7	$\mu_0 e^{\mu_1 a}$	$\lambda$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2601.0	5
M8	$\mu$	$\beta_0 e^{\beta_1 a}$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2601.3	5
M9	$\mu_0 e^{\mu_1 a}$	$\lambda_{i_i}, i = 1, 2, \dots, 6$	$s(a) = s$	2548.3	9
M10	$\mu_0 e^{\mu_1 a}$	$\beta_0 e^{\beta_1 a}$	$s(a) = s$	2621.3	5
M11	$\mu_0 e^{\mu_1 a}$	$\lambda_{i_i}, i = 1, 2, \dots, 6$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2529.5	10
M12	$\mu_0 e^{\mu_1 a}$	$\beta_0 e^{\beta_1 a}$	$\text{logit}[s(a)] = s_0 + s_1(a - \bar{a})$	2602.7	6

\* Comparison of the fit of models with different parameter specifications using the Akaike Information Criterion (AIC). A lower value of AIC indicates a better fit.

† p is the number of parameters estimated.

Poisson likelihood function. The models considered here are thus fitted to the whole genotype histories through time of the individual patients as in Sama *et al.* (2005).

The models with the specifications a(i), b(i), and c(i) above have been discussed by Sama *et al.* (2005). We now consider all the remaining possible combinations of (a), (b) and (c) (see Table 1). The models are fitted using maximum likelihood, employing a fixed order Runge Kutta method (Shampine, 1994) for the numerical integration of equation 3 and the quasi-Newton algorithm (Gill and Murray, 1976) for the maximization process. Confidence intervals were obtained by inverting the observed information matrix (Davison, 2003). The programming was implemented in Fortran 95 (*Compaq Visual Fortran Version 6.6. Compaq Computer Corporation, Houston, Texas, 2001*).

The improvement of the fit gained by considering the likelihood of a fuller model  $L_2$  containing  $p + q$  parameters with respect to the likelihood of a reduced nested model  $L_1$  containing only  $p$  parameters was compared using the likelihood ratio test. Comparison of non-nested models was done using the Akaike Information Criterion (AIC).

RESULTS

An exploratory analysis was done to assess the number of new infections gained and the proportion of existing infections that were lost. An infection present in survey X but absent in the consecutive survey X + 1 was considered as ‘Loss’ (+, -), while ‘Gain’ (-, +) was noted when an infection was present in survey X but absent in the previous survey X - 1. The total number of infections gained was calculated over all possible consecutive surveys. The total number of infections lost at survey X + 1 among those initially infected at survey X was also calculated

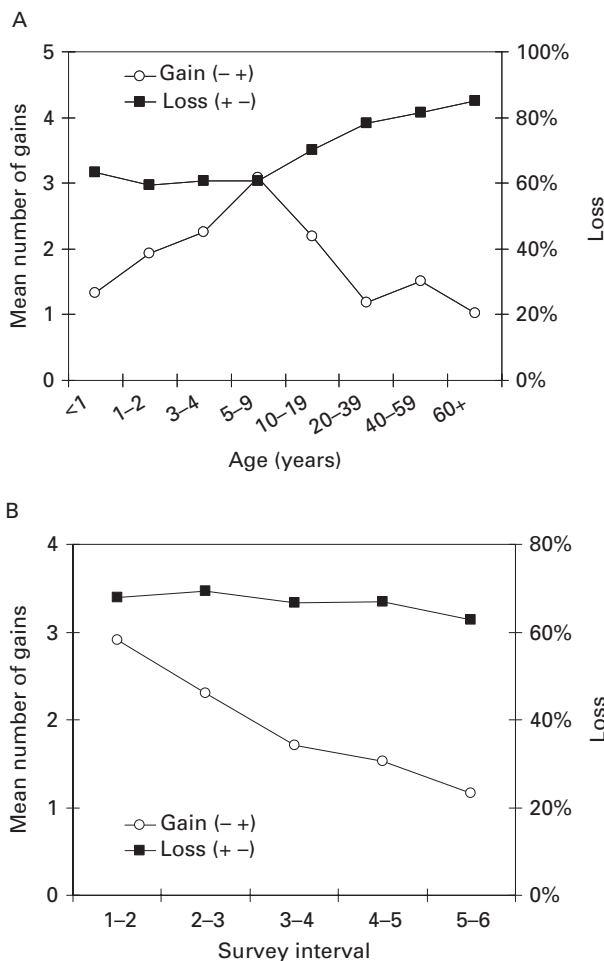


Fig. 1. (A) Mean number of newly acquired infections (Gains), and proportion of infections lost (Loss) among those initially infected, per person-interval by age group. (B) Mean number of newly acquired infections (Gains), and proportion of infections lost (Loss) among those initially infected, per person by survey interval.

over all possible consecutive surveys. The results were summarized by age group and by survey intervals. The number of new infections acquired

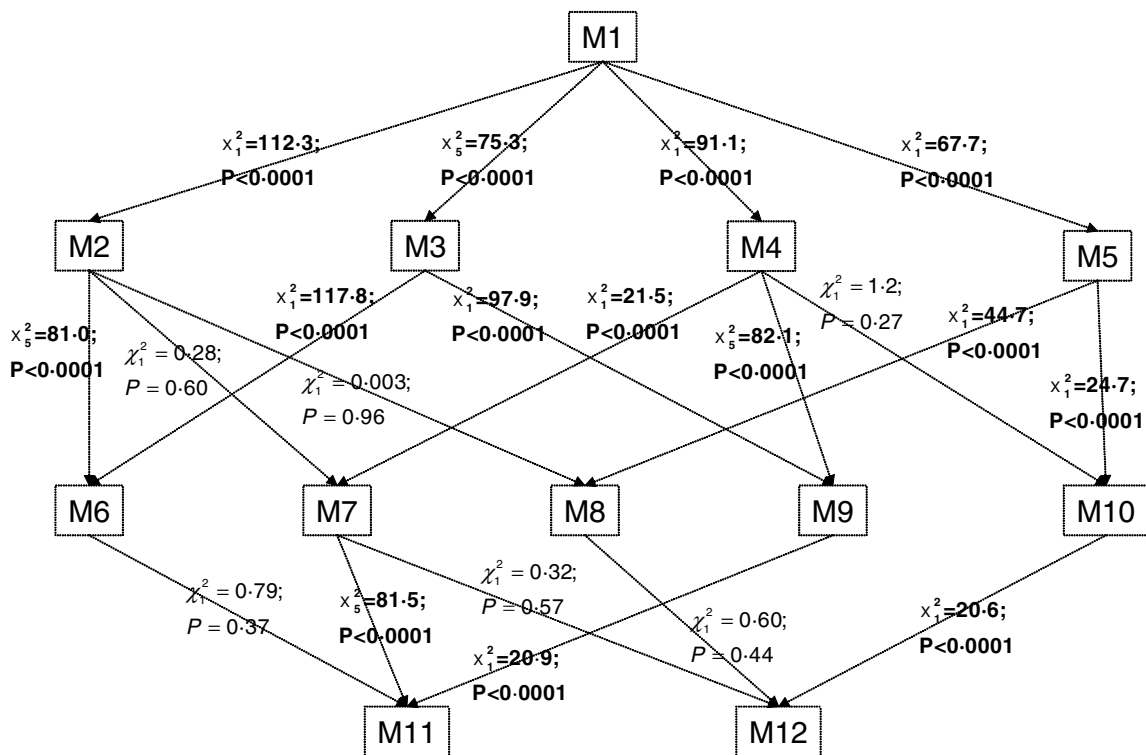


Fig. 2. Flow-chart of nested models with likelihood ratio statistics (and *P*-values) comparing the models. M1→M2: M2 is nested within M1.

increased to the age of 5-9 years and then dropped for older age groups. An increase in age of the proportion of infections lost was also observed (Fig. 1A). However, when these quantities were assessed by season, the proportion of infections lost remained fairly constant throughout the year, while there was a seasonal effect on the number of infections gained with the maximum values observed during the first 2 surveys, reflecting the high transmission rate during the wet season (Fig. 1B).

We considered a total of 12 models (Table 1). M1 is the model we reported previously (Sama *et al.* 2005) in which it was assumed that there was no age or seasonal variation in the parameters. Likelihood ratio tests indicated that all four of the extended models M2, M3, M4 and M5 in which age or seasonal effects were added, demonstrated better fits than M1 indicating that both age and seasonal variation are important (Fig. 2).

All models where detectability was expressed as a function of age (M2, M6, M7, M8, M11, M12), estimated a significant age effect, and all the estimates of  $s_1$  from these models are similar, indicating decreases in the detectability with age (Table 2). The likelihood ratio tests comparing each of these models with the corresponding reduced models with age-independent detectability all showed statistically significant differences (Fig. 2).

When the infection rate,  $\lambda$ , was assumed constant, its value (of about 17 infections gained per annum) was rather insensitive to the other parameters in the

model and somewhat higher than the rate of acquisition of apparent new infections in the raw data (Fig. 1). Although model M5 suggested that there was a strong decrease of infection rate with age, adjustment for age dependence in detectability gave an improvement in fit, and the best model (by Akaike's criterion) among those where the infection rate was allowed to vary with age was M8 (Table 1), in which there is a slight tendency for the infection rate to decrease with age (Table 2). This reflects a rather strong correlation of  $-0.78$  between the estimates of the parameters  $\lambda_1$  and  $s_1$  in model M8, indicative of moderate colinearity.

Among models where the clearance rate varies by age, the best fit was M11 (Table 1), in which there is a slight increase in clearance rate with age (Table 2). This increase was much less than that in model M4 in which both infection rate and detectability were assumed constant. The lower estimate of  $\mu_1$  in M11 reflects a rather strong positive correlation of 0.87 between the estimates of  $\mu_1$  and  $s_1$  in this model.

Neither the age trend in infection nor clearance rates were statistically significant (as measured by likelihood ratio tests comparing models M8 and M11 with their reduced models, Fig. 2). This is because, when age-dependence in the detectability was allowed for, no significant improvement in fit could be achieved by including age effects in either  $\lambda$  or  $\mu$  (compare M2 with M7 and M8 in Fig. 2; or M6 with M11) or in both of these parameters (compare M12 with M2, M7 or M8). Similarly, including age

Table 2. Parameter estimates from the different models in Table 1  
(95% Confidence Intervals in parentheses.)

Parameters	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
$\mu$	2.4 (2.2, 2.6)	2.3 (1.9, 2.6)	2.6 (2.3, 3.0)		2.4 (2.0, 2.7)	2.6 (2.2, 2.9)		2.3 (2.1, 2.5)				
$\mu_0$				1.8 (1.5, 2.1)			2.3 (1.9, 2.8)		2.0 (1.7, 2.5)	1.9 (1.5, 2.2)	2.4 (2.0, 2.8)	2.4 (2.0, 2.9)
$\mu_1$				0.020 (0.016, 0.024)			-0.003 (-0.014, 0.008)		0.020 (0.016, 0.024)	0.017 (0.011, 0.023)	0.004 (-0.005, 0.013)	-0.007 (-0.024, 0.011)
$\lambda$	16.3 (14.8, 17.8)	17.5 (15.7, 19.3)		17.1 (15.4, 18.8)			17.5 (15.7, 19.2)					
$\beta_0$					21.1 (18.6, 23.5)			17.5 (16.2, 18.8)		18.0 (15.4, 20.5)		18.0 (15.4, 20.5)
$\beta_1$					-0.015 (-0.019, -0.011)			-0.0002 (-0.007, 0.006)		-0.003 (-0.010, 0.0028)		-0.003 (-0.014, 0.008)
$s$	0.47 (0.42, 0.51)		0.49 (0.44, 0.54)	0.46 (0.42, 0.51)	0.46 (0.41, 0.51)				0.48 (0.43, 0.53)	0.47 (0.42, 0.52)		
$s_0$		-0.40 (-0.61, -0.20)				-0.33 (-0.53, -0.14)	-0.44 (-0.68, -0.20)	-0.40 (-0.58, -0.23)			-0.29 (-0.50, -0.08)	-0.45 (-0.69, -0.21)
$s_1$		-0.032 (-0.038, -0.026)				-0.035 (-0.040, -0.029)	-0.036 (-0.049, -0.022)	-0.032 (-0.041, -0.023)			-0.030 (-0.042, -0.018)	-0.036 (-0.050, -0.022)
$\lambda_1$			25.6 (20.7, 30.4)			31.1 (25.0, 37.2)			28.6 (23.2, 34.1)		31.2 (25.1, 37.2)	
$\lambda_2$			24.6 (18.8, 30.5)			22.9 (16.7, 29.2)			23.5 (17.6, 29.4)		23.0 (16.7, 29.2)	
$\lambda_3$			13.0 (7.9, 18.0)			13.8 (8.5, 19.1)			13.4 (8.3, 18.5)		13.8 (8.4, 19.1)	
$\lambda_4$			13.0 (8.5, 17.5)			13.1 (8.4, 17.8)			13.4 (8.7, 18.0)		13.2 (8.4, 17.9)	
$\lambda_5$			4.6 (1.2, 8.0)			5.4 (1.8, 9.0)			4.5 (0.9, 8.0)		5.3 (1.6, 8.9)	
$\lambda_6$			13.2 (2.5, 24.0)			21.6 (7.8, 35.5)			19.2 (5.6, 33.0)		21.0 (7.2, 34.9)	

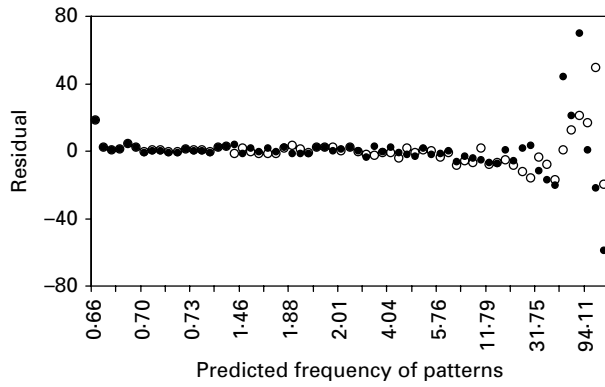


Fig. 3. Plot of the difference between the observed and predicted frequency of the 63 patterns against the predicted frequency. ● : model M1, (identical to the model of Sama *et al.* (2005)); ○ : model M6, (best fit).

dependence in both  $\lambda$  and  $\mu$  (M10) did not improve the fit significantly compared with a model with age dependence in only  $\mu$  (M4).

The overall best fitting model by Akaike’s criterion is M6, which included seasonal variation in  $\lambda$  as well as age dependence in  $s$ . Indeed, all models with seasonal patterns in the infection rate fitted better than reduced models with constant  $\lambda$ . The highest 2-month specific infection rate estimated from model M6 was 31 new infections per year, during the period June to August. The rates gradually reduce as we move towards the dry season (the low transmission period) and peaks again as we enter the wet season. This pattern is the same for all the models (M3, M6, M9, and M11) where the infection rates were allowed to vary seasonally with similar values estimated for each of the 6 different infection rates  $\lambda_i$  in the different models (Table 2).

The estimates of the parameters describing the detection process, and those measuring seasonality in infection rates were insensitive to the other variables in the model (Table 2).

The graphical fits are consistent with the results of formal evaluation of the 12 models by AIC. The frequency of the 63 observed sequences are closer to predictions when the overall best model, M6, is used than when M1 is used (Fig. 3). This is closely followed by the fits from model M11 (which is not significantly different from M6), and the fits from M3 and M9. The largest residual using M6 corresponded to sequence (010000) which occurred 147 times (expected frequency 107.8). Three other sequences had residual frequencies  $>20$  or  $<-20$ . All of these four sequences represented infections that were observed at only one survey, thus suggesting that the model fits least well for infections that persist for only a short time. When we assessed the ratio between the observed and expected frequencies, sequence (111111) which occurred 19 times, but with an expected frequency of only 0.9,

gave a very poor fit. A possible explanation for these discrepancies is that the true distribution of the durations of infections has a greater variance than is assumed by the exponential model.

Though the models are fitted to the frequency of the observed sequences and not to the observed mean multiplicity, equations 3 and 4 nevertheless predict the true numbers of distinct genotypes, as well as the numbers of distinct genotypes observed in each sample (observed mean multiplicity). There is a tendency for the observed mean multiplicity of infection to rise and then fall again as age increases. The plots of the expected multiplicity (equation 4) against age are indistinguishable for models M4, M5, M8, and M10, while that of M6 is also similar to M9, and these plots tend to follow the pattern in the observed data. The plots for the expected true multiplicity (equation 3) for models M4, M5, and M10 are indistinguishable, but differ from that of M8, while that of M6 differs from that of M9. There is a greater tendency to follow the pattern in the observed data with the plots of the expected true multiplicity using models M4, M5, M9 and M10, than for M8 and M6 that were reported as the best models above (see Fig. 4). Because of this discrepancy, we were concerned that the conclusion obtained for the clearance rate may be an artifact of the type of function considered for the clearance rate. We attempted a more flexible function by assuming that

$$\mu(a) = \mu_0 e^{\mu_1 a + \mu_2 a^2}, \quad \text{where } \mu_0 > 0, \quad -\infty < \mu_1, \mu_2 < \infty$$

This, however, gave no significant improvement of fit compared with any of models M4, M7, M10, M9, M11, and M12.

We also attempted the following 2 logistic functions for the clearance rate.

$$\mu(a) = \mu_0 + (\mu_1 - \mu_0) \frac{1}{1 + (\frac{a^*}{a})^k},$$

where  $\mu_0, \mu_1, a^* > 0, k < 0$ .

$$\mu(a) = \frac{\mu_0 \mu_1}{\mu_0 + (\mu_1 - \mu_0) e^{-\delta a}}, \quad \text{where } \mu_0, \mu_1, \delta > 0.$$

However, the parameters in all the models fitted by substituting the clearance rate in models (M4, M7, M9, M10, M11, and M12) with the above two functional forms were not identifiable.

Similarly, we attempted the following flexible functional form for the infection rate in models M5, M8, M10, and M12;

$$\lambda(a) = \beta_0 e^{\beta_1 a + \beta_2 a^2}, \quad \text{where } \beta_0 > 0, \quad -\infty < \beta_1, \beta_2 < \infty$$

The results were also not significantly different from the initial models.

We also attempted the following form for the infection rate where it is allowed to vary both by age



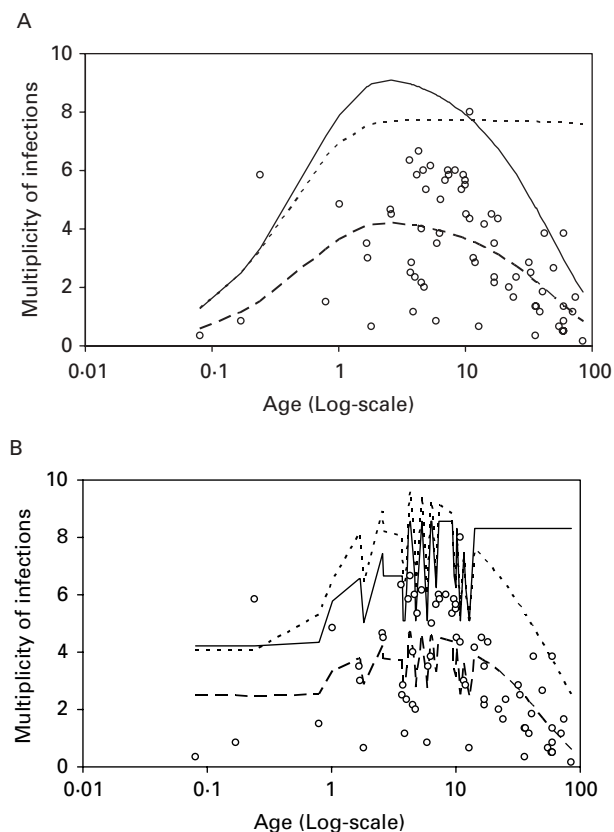


Fig. 4. (A) Mean multiplicity of infections at baseline.  $\circ$ : observed mean multiplicity;  $---$ : fit to the observed mean multiplicity ( $\tilde{n}(a, 0)$ ) using model M5;  $—$ : predicted true mean multiplicity ( $n(a, 0)$ ) using model M5;  $- - -$ : predicted true mean multiplicity ( $n(a, 0)$ ) using model M8. (B) Mean multiplicity of infections at baseline.  $\circ$ : observed mean multiplicity;  $---$ : fit to the observed mean multiplicity ( $\tilde{n}(a, 0)$ ) using model M6;  $—$ : predicted true mean multiplicity ( $n(a, 0)$ ) using model M6;  $- - -$ : predicted true mean multiplicity ( $n(a, 0)$ ) using model M9.

and season,

$$\lambda(a) = \lambda_i e^{\lambda_i a}, \quad \text{where } -\infty < \lambda < \infty, \lambda_i > 0, \\ i = 1, 2, \dots, 6$$

and evaluated models M1, M2, M4, M7 with this expression substituted for the infection rate. The model obtained from M1 by replacing the infection rate with the above expression gave better results (AIC = 2572.0) than M3 and M5 where the infection rates were allowed to vary only by age or by season. Similarly, the model obtained from M4 (AIC = 2547.4) also gave a better fit than M10 where the infection rate was allowed to vary only by age; and gave a similar fit to M9 where the infection rate was allowed to vary only by season. However, the parameters in the models obtained from M2 and M7 (which represents the most interesting case since the detectability is allowed to vary by age) were not identifiable. The models obtained from M1 and M4 indicated a seasonal variation in infection rates with

a slow decrease with age and that obtained from M4 also indicated a significant increase in clearance rate with age.

The best model by the AIC criterion remained M6.

As in our previous work (Sama *et al.* 2005) our analysis assumed that re-infection with a given genotype is a rare event. In order to assess the impact of this assumption, we excluded the most frequent genotype from the dataset, and repeated the analysis of models M1–M12. The infection and clearance rates obtained from this modification were very similar to those from the complete dataset, while detectability estimates were rather lower, (e.g. reduced to  $s = 0.38$  in model M1). The AIC results were similar to those in the full dataset.

### DISCUSSION

An important gap in knowledge of the epidemiology of malaria is in the effect of acquired immunity on the duration of infection. The rate at which *Plasmodium falciparum* infections are cleared is often thought to be highly affected by acquired immunity but the empirical evidence for this is weak. Most studies on this are based on data obtained from light microscopy, a technique which cannot distinguish between persisting and new infections. The failure to detect low-density infections and the inability to distinguish the concurrent infections within a host at a given time also potentially lead to bias in estimates of infection and recovery rates from light microscopy data (for instance as in the studies of Bekessy, Molineaux and Storey, 1976; Kitua *et al.* 1996; Sama *et al.* 2004). A single continuously infected individual, appears by microscopy neither to contribute to the pool of new infections nor to that of infections that have been cleared. Durations estimated from microscopy data will also highly depend on whether the data were derived from repeated cross-sectional studies or from longitudinal studies (Sama *et al.* 2004).

With the availability of PCR genotype specific transition data, it is now apparent that continuously infected individuals are experiencing repeated superinfection and clearance of specific genotypes, and PCR analyses can therefore give more reliable estimates of infection and recovery rates. The estimate of the overall infection rate in our study, as might have been expected, is much higher than in studies using microscopy data. Estimates of duration of infection from PCR, which are for a single monoclonal infections (Smith *et al.* 1999b; Smith and Vounatsou, 2003; Sama *et al.* 2005), tend to be shorter than those from microscopy data (Macdonald and Gökkel, 1964; Bekessy *et al.* 1976; Kitua *et al.* 1996; Sama *et al.* 2004) which are for polyclonal combinations of old and new infections.

Our previous model for *P. falciparum* genotype data from a study in Northern Ghana (Sama *et al.*

2005) assumed that the transition rates and the detection process were homogeneous. The present analysis extends this to allow different age dependences in the clearance rate, infection rate, and detectability and thus for age dependence in multiplicity of infection. Some of the models also allow for seasonal variation in the infection rates. Age dependence in multiplicity is generally found in highly endemic areas (Ntoumi *et al.* 1995; Smith *et al.* 1999c) but its relationship to acquired immunity has been unclear. No other PCR study has analysed transition rates with a complete age representation of the population.

We find that there is a decrease in detectability with increasing age. Since the parasite densities decrease with age and it is expected that a genotype is less likely to be detected if the density is low, the decrease with age in detectability was anticipated. Detectability varied from about 60% in younger individuals to 10% in adults.

Immunity is known to increase with age and it has often been assumed that duration of infection must decrease as immunity is acquired. Some studies have found that infections are of very short duration in infants (Walton, 1947; Kitua *et al.* 1996; Smith *et al.* 1999b) and one study in young children estimated that duration increases with age (Smith *et al.* 2003). In our analysis, the models that did not include age dependence in detectability estimated significant increases in the clearance rate with age. However, our best fitting models allow for age dependence in detectability and found no residual effect of age on clearance rates, but we have to treat this conclusion with caution because the parameter estimates for the age effect on clearance are rather strongly correlated with those for age dependence in detectability.

The estimated duration of infection was not greatly modified by allowing for seasonal variation in the infection process, although infections were found to be highly seasonal. The use of 6 different parameters for the infection rates allowed us to obtain season-specific estimates for the infection rates with the highest infection rate estimated from our best model, M6, is 31 new infections per year, during the period June to August which is the peak transmission season in this area (Binka *et al.* 1994; Baird *et al.* 2002). This implies that individuals may acquire as many as 5 new infections during this period.

In separate models (M5, M8, M10, M12) we allowed the infection rate to vary only by age, the best model among these was M8 that indicated a decrease in infection rate with age, but this was not statistically significant (see Fig. 2 for comparison of M8 and M2). Some studies have suggested that infection rates in children are higher than in older people (Rogier and Trape, 1995), as would be expected if immunity to infection is acquired. However, the attractiveness of the human host to mosquitoes in general increases with host size (Port, Boreham and Bryan, 1980),

which should lead to an increase in infection rates with age (Smith *et al.* 2004). The overall best model among the 12 models we considered was M6 that included season- but not age dependence in infection rates.

Our results should be interpreted cautiously because none of our models allows simultaneously for seasonal variation in infection rates or detectability, variation of infection rates with age, or variation of clearance rate with age. Further work needs to be done to capture all these variations and this will need a substantial amount of information, covering a wide range of age groups, from different endemic, geographical, and cultural settings. The present models also ignore heterogeneity between genotypes, which may also be important.

We assumed age to be the only factor leading to heterogeneity in infection rates across the population. However, on 26% of visits respondents indicated that they slept under bednets. It is likely that these nets moderated the infection rate but had little effect on the clearance rate. Patterns of treatment for febrile illness in the KND are complex (Owusu *et al.* manuscript submitted) often involving ineffective medications or inappropriate dosing. We did not record treatments used by the participants during the study or attempt to model effects of treatment but it is clear that the use of anti-malarial treatments will increase the clearance rate and so may have biased downwards our estimates of duration.

A further improvement would be to directly incorporate information about parasite densities. The clearest effect of acquired immunity is in reducing the mean parasite density, and this must account for a large element of the decrease in detectability with age. The question of how to incorporate the extra information contributed by the parasite density while avoiding identifiability problems deserves further attention. There is likely to be a change in both detectability and clearance rate if this information is well accounted for.

We thank the staff of the Navrongo Research Health Centre involved in the field work and the villagers for their co-operation. We also thank Beatrice Glinz for her contribution to the genotyping. Wilson Sama was in receipt of stipends from the Stipendiumkommission of the Amt für Ausbildungsbeiträge of the Canton of Basel, and from the Swiss National Science Foundation (Project no. 3300CO-105994/1).

#### REFERENCES

- Aron, J. L. (1988). Mathematical-modeling of immunity to malaria. *Mathematical Biosciences* **90**, 385–396.
- Baird, J. K., Owusu-Agyei, S., Utz, G. C., Koram, K., Barcus, M. J., Jones, T. R., Fryauff, D. J., Binka, F. N., Hoffman, S. L. and Nkrumah, F. N. (2002). Seasonal malaria attack rates in infants and young children in northern Ghana. *American Journal of Tropical Medicine and Hygiene* **66**, 280–286.



- Bekessy, A., Molineaux, L. and Storey, J.** (1976). Estimation of incidence and recovery rates of *Plasmodium falciparum* parasitaemia from longitudinal data. *Bulletin of the World Health Organisation* **54**, 685–691.
- Binka, F. N., Morris, S. S., Ross, D. A., Arthur, P. and Aryeetey, M. E.** (1994). Patterns of malaria morbidity and mortality in children in northern Ghana. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **88**, 381–385.
- Davison, A. C.** (2003). *Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK.
- Dietz, K., Molineaux, L. and Thomas, A.** (1974). A malaria model tested in the African savannah. *Bulletin of the World Health Organisation* **50**, 347–357.
- Felger, I., Irion, A., Steiger, S. and Beck, H. P.** (1999). Genotypes of merozoite surface protein 2 of *Plasmodium falciparum* in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93** (Suppl. 1), 3–9.
- Gill, P. E. and Murray, W.** (1976). *Minimization Subject to Bounds on the Variables*. National Physical Library, NPL Report NAC 72.
- Kitua, A. Y., Smith, T., Alonso, P. L., Masanja, H., Urassa, H., Menendez, C., Kimario, J. and Tanner, M.** (1996). *Plasmodium falciparum* malaria in the first year of life in an area of intense and perennial transmission. *Tropical Medicine and International Health* **4**, 475–484.
- Koram, K. A., Owusu-Agyei, S., Fryauff, D. J., Anto, F., Atuguba, F., Hodgson, A., Hoffman, S. L. and Nkrumah, F. K.** (2003). Seasonal profiles of malaria infection, anaemia, and bednet use among age groups and communities in northern Ghana. *Tropical Medicine and International Health* **8**, 793–802.
- Macdonald, G.** (1950). The analysis of infection rates in diseases in which superinfection occurs. *Tropical Diseases Bulletin* **47**, 907–915.
- Macdonald, G. and Göckel, G. W.** (1964). The malaria parasite rate and interruption of transmission. *Bulletin of the World Health Organization* **31**, 365–377.
- Ntoumi, F., Contamin, H., Rogier, C., Bonnefoy, S., Trape, J. F. and Mercereau-Pujalon, O.** (1995). Age-dependent carriage of multiple *Plasmodium falciparum* merozoite surface antigen-2 alleles in asymptomatic malaria infections. *American Journal of Tropical Medicine and Hygiene* **52**, 81–88.
- Port, G. R., Boreham, P. F. L. and Bryan, J. H.** (1980). The relationship of host size to feeding by mosquitoes of the *Anopheles gambiae* Giles complex (Diptera: Culicidae). *Bulletin of Entomological Research* **70**, 133–144.
- Rogier, C. and Trape, J. F.** (1995). Study of premunition development in holo- and meso-endemic malaria areas in Dielmo and Ndiop (Senegal): preliminary results, 1990–1994 [Article in French]. *Médecine Tropicale (Mars)* **55**, (Suppl.) 71–76.
- Sama, W., Killeen, G. and Smith, T.** (2004). Estimating the duration of *Plasmodium falciparum* infection from trials of indoor residual spraying. *American Journal of Tropical Medicine and Hygiene* **70**, 625–634.
- Sama, W., Owusu-Agyei, S., Felger, I., Vounatsou, P. and Smith, T.** (2005). An immigration-death model to estimate the duration of malaria infection when detectability of the parasite is imperfect. *Statistics in Medicine* (in the Press).
- Shampine, L. F.** (1994). *Numerical Solution of Ordinary Differential Equations*, Chapman and Hall, London.
- Smith, T., Felger, I., Fraser-Hurt, N. and Beck, H. P.** (1999a). Effect of insecticide-treated bed nets on the dynamics of multiple *Plasmodium falciparum* infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93** (Suppl. 1), 53–57.
- Smith, T., Felger, I., Kitua, A., Tanner, M. and Beck, H. P.** (1999b). Dynamics of multiple *Plasmodium falciparum* infections in infants in a highly endemic area of Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93** (Suppl. 1), 35–39.
- Smith, T., Beck, H.-P., Kitua, A., Mwangusye, S., Felger, I., Fraser-Hurt, N., Irion, A., Alonso, P., Teuscher, T. and Tanner, M.** (1999c). Age dependence of the multiplicity of *Plasmodium falciparum* infections and of other malariological indices in an area of high endemicity. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **93** (Suppl. 1), 15–20.
- Smith, T. and Vounatsou, P.** (2003). Estimation of infection and recovery rates for highly polymorphic parasites when detectability is imperfect, using hidden Markov models. *Statistics in Medicine* **22**, 1709–1724.
- Smith, T., Killeen, G., Lengeler, C. & Tanner, M.** (2004). Relationships between the outcome of *Plasmodium falciparum* infection and the intensity of transmission in Africa. *American Journal of Tropical Medicine and Hygiene* **70** (Suppl. 2), 80–86.
- Walton, G. A.** (1947). On the control of malaria in Freetown, Sierra Leone. 1. *Plasmodium falciparum* and *Anopheles gambiae* in relation to malaria occurring in infants. *Annals of Tropical Medicine and Parasitology* **41**, 380–407.