# Bayesian age-stage modelling of *Plasmodium falciparum* sequestered parasite loads in severe malaria patients

T. SMITH[1]*, K. DIETZ[2], P. VOUNATSOU[1], I. MÜLLER[1]†, M. ENGLISH[3] *and* K. MARSH[3]

[1] *Swiss Tropical Institute, Socinstrasse 57, Postfach, CH-4002, Basel, Switzerland*
[2] *Institut für Medizinsche Biometrie, Westbahnhofstrasse 55, D-72070, Tübingen, Germany*
[3] *Kenya Medical Research Institute Laboratories (KEMRI)/Wellcome Trust Kilifi District Hospital, PO Box 230, Kilifi, Kenya*

(*Received 12 January 2004; revised 1 February 2004; accepted 2 March 2004*)

SUMMARY

A discrete-time age-stage model is proposed for estimating the number of sequestered parasites in severe malaria patients. A Bayesian Markov chain Monte Carlo (MCMC) approach is used to model the dynamics of *Plasmodium falciparum* parasitaemia in 107 paediatric patients in a randomized controlled trial of quinine and artemether in Kenya, in whom 4-hourly peripheral parasitaemia determinations were made. The MCMC approach allows the model to be fitted simultaneously to the entire dataset, providing point and interval estimates for both population and individual patient parameters. Analysis of a simulated dataset indicated that the models gave good estimates of the distribution of parasites between different stages on enrolment, for patients with a wide range of initial states. The analysis of the Kenyan patients suggested that there is considerable variation between patients within the same centre, in both the proportion of sequestered parasites and the intrinsic rate of increase of the parasite population in the absence of treatment. The resulting models should prove a useful tool for cross-validating biochemical approaches for estimating the sequestered load.

Key words: sequestration, mathematical model, Kenya.

## INTRODUCTION

The pathology of *Plasmodium falciparum* malaria is associated with the adherence of late-stage trophozoites and schizonts to endothelia (e.g. Warrell *et al.* 1990; White & Ho, 1992) and with the release of toxins when schizonts rupture. However, diagnosis of malaria continues to rely mainly on microscopy of peripheral blood smears and there are no methods that can be routinely applied in a clinical setting for estimating the numbers of sequestered parasites, or of rupturing schizonts in a living host.

One approach for estimating the sequestered load is to fit a model to time-series of peripheral parasitaemia assessments. The theoretical dynamics of peripheral parasite densities in untreated individuals are well known (e.g. White, Chapman & Watt, 1992; Fig. 1a). However, data to fit such models are available only from (i) serendipitous clinical trials where a treatment failed (e.g. the 4 patients on ciprofloxacin studied by White *et al.* (1992); (ii) studies of asymptomatic individuals: these include experimental infections in the pre-clinical growth phase (e.g. Fairley, 1947; Cheng *et al.* 1997); (iii) semi-immune individuals where treatment is not indicated; (iv) or historical data from neurosyphilis patients (e.g. Collins & Jeffery, 1999).

Severe malaria patients, with closely monitored parasitaemia are the richest source of longitudinal parasitological data with which to estimate sequestered loads. Moreover, the estimation of the sequestered load and the degree of parasite synchronization in such patients could provide important markers of severity since disease severity is thought to be related more to the sequestered than circulating parasite load (White & Ho, 1992; Gravenor, van Hensbroek & Kwiatkowski, 1998; Gravenor *et al.* 2002). Indeed, Gravenor *et al.* (1998, 2002) fitted statistical models by maximum likelihood to sequential determinations of parasite densities in a trial of quinine and artemether in The Gambia and found a relationship between the estimated number of sequestered parasites at the start of treatment and the severity of the disease. The models were also consistent with the *in vitro* findings that quinine is relatively effective in clearing young ring stages and trophozoites, while artemether is also highly active against schizonts (ter Kuile *et al.* 1993).

We have now implemented an alternative discrete-time approach based on 4-h stages. We have fitted this to 4 hourly peripheral parasitaemia determinations from 107 severe malaria patients treated with

* Corresponding author: Swiss Tropical Institute, Socinstrasse 57, Postfach, CH-4002, Basel, Switzerland. Tel: +41 61 284 8273. Fax: +41 61 271 7951. E-mail: Thomas-A.Smith@unibas.ch
† Current address: Papua New Guinea Institute of Medical Research, PO Box 60, Goroka, EHP, Papua New Guinea.

*Parasitology* (2004), **129**, 289–299. 
DOI: 10.1017/S003118200400575X Printed in the United Kingdom

Table 1. Measured and estimated characteristics of patients

(Figures in square brackets are the standard deviations.)

| | Quinine | Artemether |
|---|---|---|
| Total patients included | 53 | 54 |
| Mean $Log_{10}$ (peripheral parasite density at baseline) | 4·81 [0·89] | 4·99 [0·93] |
| Mean number of slides | 14·8 [3·3] | 14·1 [2·8] |
| Number of positive slides | 10·9 [3·3] | 9·5 [3·0] |
| Estimates | | |
| Mean $Log_{10}$ (total parasite density at baseline) | 5·19 [0·84] | 5·28 [0·78] |
| Median Proportion of parasites sequestered at baseline (range) | 0·44 (0·04, 0·91) | 0·30 (0·02, 0·82) |

either quinine or artemether in a randomized controlled trial in Kilifi, Kenya (Murphy *et al*. 1996). We use a Bayesian Markov Chain Monte Carlo (MCMC) approach to fit the model, which makes it feasible to estimate some parameters at the level of the individual patient, but within the same fitting process to make population level estimates of other parameters. The uncertainties in the population model are thus taken into account in the estimates of the individual-level parameters. By assuming constant sojourn times in each parasite stage, rather than modelling the flow-rates as proportional to the number of parasites in the stage, the model attains greater realism and generally gives more plausible estimates of total parasite loads than does that of Gravenor *et al*. (2002). It allows both for stage specificity of drug action and for time delays before drugs become active.

## MATERIALS AND METHODS

### *Patients*

The statistical models have been fitted to sequential parasite density data from the trial of Murphy *et al*. (1996). This was a randomized controlled trial of artemether and quinine for the treatment of cerebral malaria carried out in Kilifi, Kenya. Children admitted to the hospital with coma and *P. falciparum* parasitaemia were treated with either intramuscular artemether (3·2 mg/kg loading dose followed by 1·6 mg/kg daily) or intravenous quinine (20 mg/kg loading dose followed by 10 mg/kg every 8 h). In total, 160 children satisfied the admission criteria of the trial. Parasite densities in this trial were assessed every 4 h until clearance. The present analyses considered the data from a total of 107 such patients, from whom at least 8 such 4 hourly parasite density determinations were available (Table 1). Of these patients, 53 were treated with quinine and 54 with artemether.

### *Model of parasite dynamics*

We divide the 48-h cycle of blood-stage *P. falciparum* malaria into 4-hourly intervals indexed by $a = 1 \ldots 12$ and denote as $N_{k,a,t}$ the expected number of parasites in patient $k$, at the $a$th stage, at 4-hourly time-point $t$. We denote the probability of a parasite in patient $k$, stage $a$ surviving from $t$ to $t+1$ as $S_{k,a,t}$. It follows that:

$$N_{k,a,t} = S_{k,a-\mathbf{1},t-\mathbf{1}} N_{k,a-\mathbf{1},t-\mathbf{1}} \quad \text{for all } a\text{: } 1 < a \leqslant 12 \tag{1.1}$$

and

$$N_{k,\mathbf{1},t} = R_k S_{k,\mathbf{12},t-\mathbf{1}} N_{k,\mathbf{12},t-\mathbf{1}}, \tag{1.2}$$

where $R_k$ is the intrinsic rate of increase per asexual cycle. Using these recursive relationships the expected numbers of parasites in each stage at each time-point can be determined, conditional on estimates of $S_{k,i,t}$, $R_k$, and the number of parasites at each stage at baseline, $N_{k,a,\mathbf{0}}$. It follows that the parasites at each 4-hourly stage, at each time-point, are treated as the survivors and progeny of a cohort of parasites in a given stage at baseline (Fig. 1).

### *Specification of intrinsic multiplication rate and stage-specific baseline densities*

Since severe malaria is indicative of a failure to control parasites, at the start of treatment we expect the parasite population to be growing. If the parasites were completely unsynchronized, and clearance of infected erythrocytes negligible, this would mean that the numbers of parasites at the different stages would on average (considering the whole population of patients, and remembering that we divide the erythrocytic cycle into twelve 4-hourly intervals) be in the ratio:

$$N_{k,a,\mathbf{0}} = N_{k,\mathbf{1},\mathbf{0}} \left( \frac{11R_k + 1}{12R_k} \right)^{a-\mathbf{1}}. \tag{1.3}$$

(For details of the derivation of this formula see the Appendix.)

To model deviations from these relative proportions, the stage distributions of the parasites

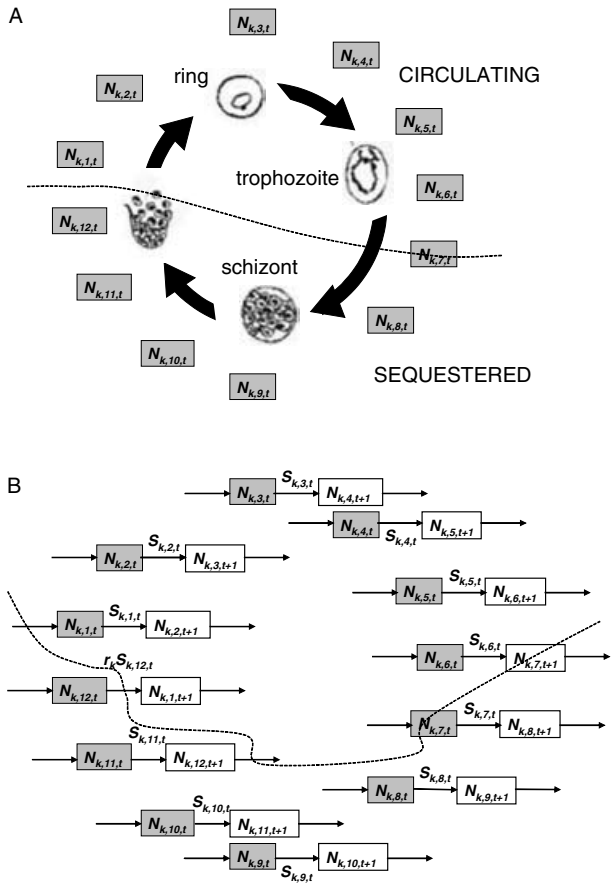


Fig. 1. Asexual cycle of *Plasmodium falciparum* and structure of the model. (A) Asexual cycle showing notation used for parasite densities $N_{k,a,t}$ is the expected density of parasites at stage $a$ at time $t$ in patient $k$. Note that the dotted line delimiting peripheral and sequestered parasites intentionally cuts the boxes with index 7, since sequestration is assumed to occur while parasites are in this stage. (B) Model for change in numbers of parasites over one 4-h interval. The open boxes give the expected numbers of parasites at time $t+1$. The term adjacent to the arrows is the multiplying factor that determines how these expected numbers are related to each other.

within each patient at baseline were modelled using a circular function, $f_k(a)$, based on the von Mises distribution (Evans, Hastings & Peacock, 2000), comprising an unsynchronized (stage-independent) component ($\beta_k$), and a synchronized component, where:

$$f_k(a) = \exp(\kappa \cos(\phi - \theta_k)) + \beta_k, \tag{1.4}$$

$\phi = 2\pi a/12$, $\theta_k$ specifies the phase for patient $k$, and $\kappa$ determines the width of the peaks. To obtain a proper distribution for the proportions of parasites in each stage, we can divide the product $N_{k,a,0}\, f_k(a)$ by the sum over all stages. To convert these proportions to absolute densities we multiply by the total parasite load at baseline, $T_k = \sum_a N_{k,a,0}$ which is therefore a further parameter of the model i.e.:

$$N_{k,a,0} = T_k \frac{\left(\frac{11R_k+1}{12R_k}\right)^{a-1} f_k(a)}{\sum_a \left[\left(\frac{11R_k+1}{12R_k}\right)^{a-1} f_k(a)\right]}. \tag{1.5}$$

Thus, 4 patient-specific parameters, $T_k$, $\beta_k$, $\theta_k$, and $R_k$ are used for each patient (in addition to $\kappa$, which is assigned at the level of the patient population) to specify the distribution of the parasites between stages at the start of treatment (see Table 2).

### *Pharmacodynamics*

To model drug-induced parasite death we assume a separate function measuring the stage-specific effect on survival of each of the two drugs ($d = 1, 2$), basing this on published estimates of the relative inhibition of [$^3$H] hypoxanthine incorporation, determined *in vitro* (ter Kuile *et al.* 1993) (see Fig. 2A). For each drug we estimate one parameter, $\mathrm{M}_d$, corresponding to the maximal inhibition, so that for the most susceptible stage of parasite, in the presence of maximally effective drug concentrations:

$$S_{k,a,t} = 1 - \mathrm{M}_d. \tag{1.6}$$

Denoting the relative inhibition by drug $d$ of parasites of stage $a$ as $\rho_{d,a}$ (where $\rho_{d,a}$ varies between 0 and 1), the survival of parasites at other stages in the cycle is:

$$S_{k,a,t} = 1 - \mathrm{M}_d \rho_{d,a}. \tag{1.7}$$

To allow for the initial delay before drug effects are seen, we assume the same kinetics as those of the inhibition of [$^3$H] hypoxanthine incorporation determined *in vitro* (ter Kuile *et al.* 1993), and specify a quantity $\omega_t$ corresponding to the proportion, at time $t$, of the maximal inhibitory effect of the drug (see Fig. 2B). Thus, in general, parasite survival is given by:

$$S_{k,a,t} = 1 - \mathrm{M}_d \rho_{d,a} \omega_t. \tag{1.8}$$

### *Fitting the model to peripheral parasite densities*

It follows that, in the reference model the expected number of parasites at stage $a$, patient $k$ and time-point $t$, $N_{k,a,t}$, is a function of quantities $\rho_{d,a}$ and $\omega_t$ taken from ter Kuile *et al.* (1993) and the parameters $T_k$, $\beta_k$, $\theta_k$, $\mathrm{M}_d$, $R_k$, and $\kappa$ to be estimated.

Following White *et al.* (1992) we assume that parasites are sequestered for rather less than half of the duration of the 48-h cycle and that those in the

Table 2. Prior and posterior distributions of parameters

| Symbol | Interpretation | Prior distribution | Point Est. | 95% confidence limits Lower | Upper |
|---|---|---|---|---|---|
| Population level estimates | | | | | |
| $M_1$ | Maximum parasitocidal effect of quinine | Normal | 0·76 | 0·73 | 0·79 |
| $M_2$ | Maximum parasitocidal effect of artemether | Normal | 0·78 | 0·76 | 0·81 |
| $\sigma^2_{e,1}$ | Error variance (quinine) | Inverse gamma | 0·00024 | 0·00022 | 0·00027 |
| $\sigma^2_{e,2}$ | Error variance (artemether) | Inverse gamma | 0·00028 | 0·00025 | 0·00032 |
| $\kappa$ | Width of the peak in the stage distribution of the synchronized subpopulation of parasites | Invariant | 10 | — | — |
| Patient level parameters | | | | | |
| $\theta_k$ | Location parameter for peak | Uniform (0, $2\pi$) | Separate estimates of these parameters were made for each patient | | |
| $R_k$ | Intrinsic multiplication rate | Uniform (1, 16) | | | |
| Random effects* | | | | | |
| Prior for patient-level parameters | | | | | |
| $T_k$ | Total parasite load at baseline | $\ln(T_k) \sim \text{Normal}(\varsigma, \sigma^2_T)$ | 11·9 | 11·5 | 12·3 |
| $\beta_k$ | Stage independent contribution to parasite load at baseline | $\ln(\beta_k) \sim \text{Normal}(\xi, \sigma^2_b)$ | 7·5 | 7·0 | 8·0 |

* The parameters for the different patients were assumed to be related via distributions of patient-level effects parameterized via the hyperparameters, $\varsigma$, $\sigma^2_T$, $\xi$, and $\sigma^2_b$. The tabulated means and interval estimates are for the hyperparameters, $\varsigma$ and $\xi$, both of which had normal prior distributions.

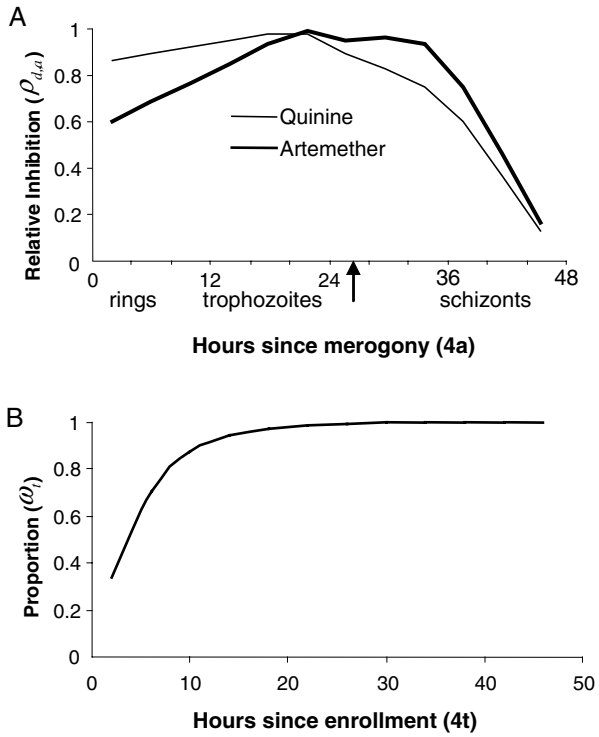


Fig. 2. Stage- and time-dependence of drug effects. (A) Relative inhibition by stage of parasite. (B) Proportion of maximal effect of drug, by observation time. Pharmacokinetics and pharmacodynamics inferred from the results of ter Kuile *et al.* (1993). The parameters used in the model correspond to 4-h stage- or time-periods.

first six 4-h stages, and half those in the 24 to 28-h stage are in the peripheral circulation. The expected number of parasites observed in 1 $\mu$l of peripheral blood at any time-point $t$, $P_{k,t}$ is therefore given by:

$$P_{k,t} = 0{\cdot}5N_{k,7,t} + \sum_{i=1}^{6} N_{k,i,t}. \quad (1.9)$$

We specify prior distributions for each of the parameters as indicated in Table 2.

To achieve approximate homogeneity of variances, we used a square-root transformation of both the observed and predicted densities ($X_{k,t}$ and $P_{k,t}$ respectively) and estimate the model using WinBugs v1.4 (Spiegelhalter *et al.* 2000), assuming normal errors. To provide estimates of the fit of the model, separately for the two drugs, the error variance, $\sigma^2_d$, was allowed to differ between quinine-treated patients ($d=1$) and those treated with artemether ($d=2$), i.e.

$$\sqrt{X_{k,t}} \sim \text{Normal}(\sqrt{P_{k,t}}, \sigma^2_d). \quad (1.10)$$

The program uses the Metropolis-Hastings algorithm to obtain posterior distributions for the parameters. Convergence was monitored using the program CODA (Best, Cowles & Vines, 1996) using Heidelberger and Welch tests (Heidelberger & Welch, 1983).

The models were also evaluated in terms of whether the estimated overall parasite density was plausible. We consider a density of approximately $2 \times 10^6$ parasites per $\mu$l to represent an approximate upper limit consistent with the patient still being alive, and therefore reject any models that give estimates outside this range for many of the patients.

*Simulations*

In addition to the analysis of the patients from the study of Murphy *et al*. (1996), a simulated dataset was analysed to evaluate the performance of the model when $N_{k,a,0}$ is known. This dataset, like that of Murphy *et al*. (1996), comprised 54 series simulated by assuming the pharmacodynamics to follow those of quinine, and 53 assuming the pharmacodynamics of artemether. A subset for each drug, covering the full range of $R_k$ from 1 to 16 were simulated assuming the parasites to be completely unsynchronized (i.e. following equation (1.3)). Other simulated patients had highly synchronized parasite populations, or densities randomly assigned to the 12 stage categories. The parasitological monitoring of the simulated dataset was assumed to be equivalent to that of Murphy *et al*. (1996).

## RESULTS

The patients included in the analysis for the two drugs had similar parasite densities at baseline, and similar follow-up periods (Table 1). However, the requirement that at least 8 slides could be analysed introduced a small bias due to exclusion of relatively low density artemether patients whose parasitaemia had cleared very quickly.

We initially attempted to simultaneously estimate $T_k$, $\beta_k$, $\theta_k$, $M_d$, $R_k$, and $\kappa$ for all 107 patients. All the parameters proved to be identifiable except for $\kappa$. When this parameter was given a non-informative prior, the estimate tended towards very high values, corresponding to the limiting case of highly synchronized parasite populations. Such models gave implausibly high total parasite load ($T_k$) estimates for many of the patients and formal convergence criteria could not be satisfied because of numerical overflows during the fitting process. Models with moderately high values of $\kappa$, fitted better than those with lower values, and so the remaining parameters were assigned conditional on a moderately high value of $\kappa = 10{\cdot}0$.

This model gave an excellent fit to the data for most of the 107 patients, despite widely differing observed clearance rates of peripheral parasites (Table 1, Fig. 3). The distributions of the estimates of the intrinsic rate of increase, $R_k$, were similar for the two drugs (Fig. 4D), in both cases with $R_k$ values covering the whole range allowed by the prior distribution. While there was thus considerable variation between values of $R_k$ for different patients, many of the individual values were estimated relatively precisely (see the confidence intervals given on Fig. 3).

We use the coefficient of variation of the densities of parasites in each of the 12 stage categories as a measure of the degree of synchronization. This was calculated as the standard deviation of the estimated density across the 12 classes, divided by the arithmetic mean density, and varied from 0·14 to 2·10. The stage distributions of the parasites within each patient at baseline were modelled as a mixture of two components: a highly synchronized component (corresponding to the peaks in the right hand panels of Fig. 3) and an unsynchronized component, corresponding to the relatively flat bases of the stage distributions. In patients with relatively unsynchronized parasites with a low coefficient of variation, most of the parasite density is contributed by the base (e.g. Fig. 3F) while in patients estimated to have highly synchronized parasites most of the density is contributed by the peak (Fig. 3A and D). There was considerable variation in the coefficient of variation, corresponding to variation between patients in the relative contributions of the two components of the density distribution (Fig. 5C).

Where the intrinsic rate of increase, $R_k$, is estimated to be much greater than 1, corresponding to an initially rapidly growing parasite population, there is a clear trend for the estimates of the ring-stage parasite densities to be higher than those for the schizonts (e.g. Fig. 3C). Conversely, where $R_k$ is close to 1, the estimated number of parasites in the base component of the distribution is similar for all the 4-h stage classes.

In patients with very rapid initial decreases in density, most of the parasites were estimated to be in the ring or trophozoite stages at baseline (Fig. 3A and E). Contrasting with this, initial increases in parasite densities observed for many of the patients could be explained by high initial sequestered loads (Fig. 3B and C). In most cases the maximum in the distribution of parasite stages at baseline was in the schizont stages, with relatively few patients estimated to have predominantly trophozoites (Fig. 4).

Almost all estimates of the total parasite density were within a reasonable range (Table 1, Fig. 5A). The exceptions were 4 patients estimated to harbour more than $2 \times 10^6$ parasites per $\mu$l at baseline. The highest density was estimated to be for Patient 7 (Fig. 3D), where the model estimates of total parasite load exceeded $4 \times 10^6$ parasites per $\mu$l.

The fit of the model, as measured by the residual variance, $\sigma^2_{e,d}$, was similar, but somewhat better (i.e. low $\sigma^2_{e,d}$) for artemether patients than for those on quinine (Table 2). Since drug allocation was randomized, we expect the parasitological status at baseline for the two groups to be comparable, and indeed, the estimates of baseline total parasite load were similar for the two drugs (Fig. 5A, Table 1). However, the estimated sequestered loads were slightly higher for patients on artemether than for quinine (Fig. 5B, Table 1), and the artemether-treated patients included more in which the parasites were estimated to be highly synchronized (high coefficient of variation) (Fig. 5C). This may reflect the small selection bias

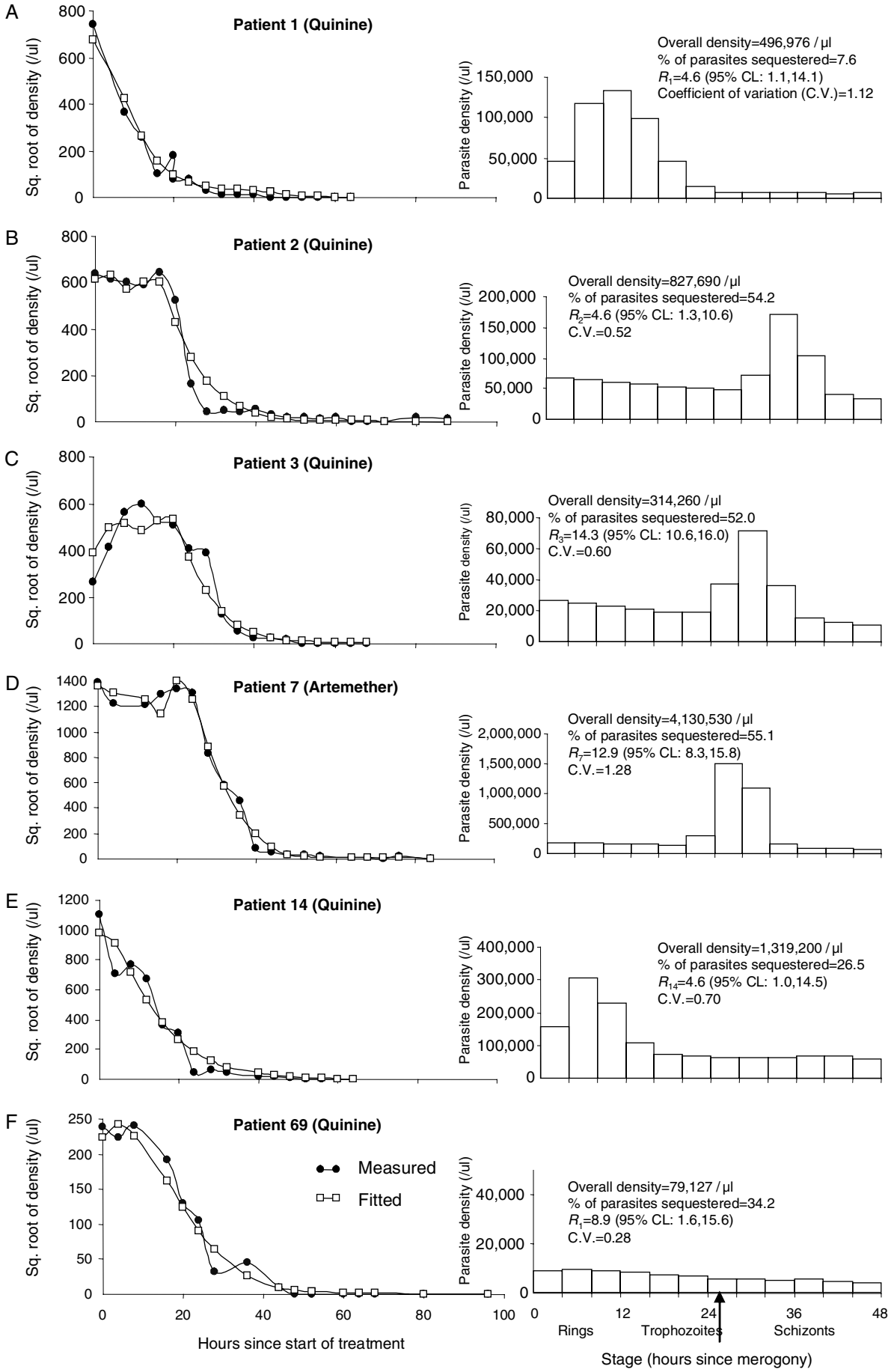


Fig. 3. For legend see opposite page.

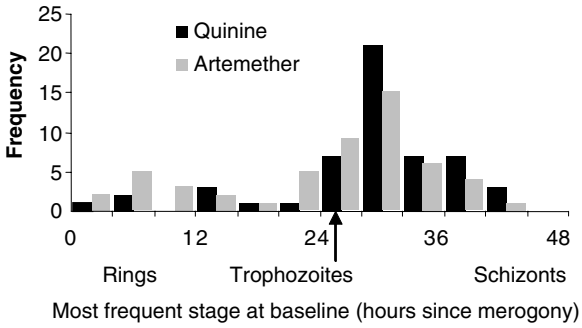


Fig. 4. Estimates of most frequent parasite stage at baseline.

resulting from the requirement that at least 8 slides were included per patient.

With simulated datasets where the parasites were unsynchronized the model gave estimates of $N_{k,a,\mathbf{0}}$ and of $R_k$ very close to the true values (Fig. 6A). When the simulated parasite population was highly synchronized the performance in terms of estimation of $R_k$ was more variable (Fig. 6B, C, and D). However, the estimated values of $N_{k,a,\mathbf{0}}$ were reasonable approximations to the true baseline status, even when the true distribution of $N_{k,a,\mathbf{0}}$ between stages was an arbitrary multimodal function (Fig. 6E).

To evaluate the extent of collinearity in the estimation of the 4 patient-specific parameters we computed the Pearson correlation coefficients between the values of $\ln(T_k)$, $\ln(\beta_k)$, $R_k$, and $\theta_k$. The absolute value of none of these correlations exceeded a value of 0·3, indicating an acceptably low level of association.

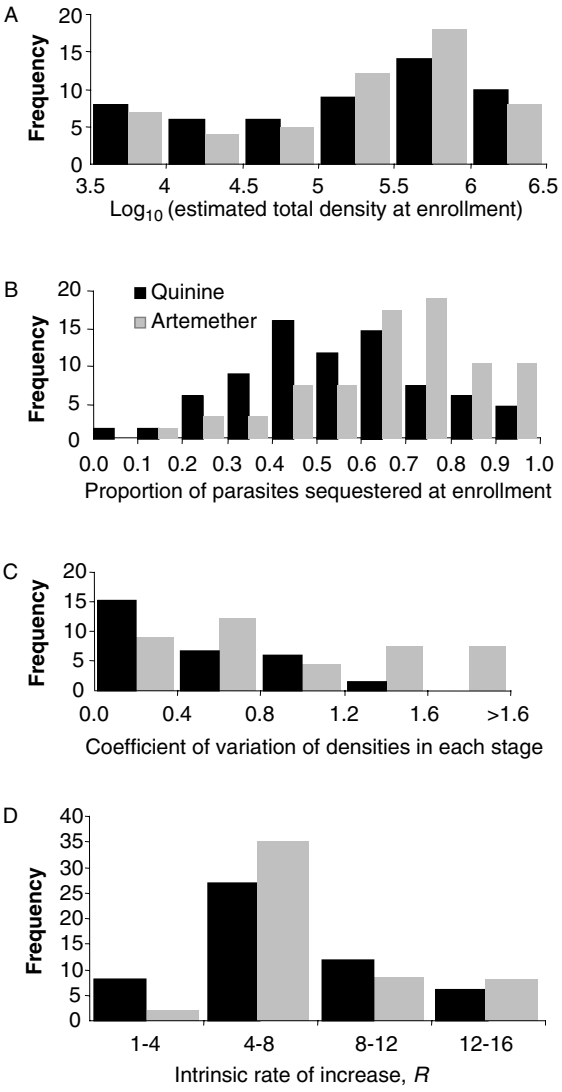


Fig. 5. Distributions of baseline characteristics estimated from the model. (A) Total parasite load. (B) Proportion of parasites sequestered. (C) Degree of synchronization. (D) Intrinsic rate of increase.

## DISCUSSION

The statistical estimation of the *in vivo* distribution of *P. falciparum* parasites between different stages of the asexual cycle represents a major challenge. *A priori* we have little information about the distribution of parasites between stages at the start of treatment, and a flexible family of curves is thus needed to describe the initial state. A consequence is that realistic models need many parameters, some of them like $\sigma_d^2$ and $M_d$ determined at the level of the patient population, while others, such as $R_k$ and those describing the baseline distribution of parasites between stages, vary between individuals.

It is often difficult to find maximum likelihood solutions for highly parameterized non-linear models, because of the difficulty of finding the maximum in a high dimension parameter space. One approach, used by Gravenor *et al*. (1998) for modelling parasite dynamics, is to fit the model separately to the data for each patient, however chance relationships within the data mean that such models do not converge for all the patients. Moreover, population-level parameters must be determined in a distinct step.

In contrast, the Bayesian MCMC approach we use can readily estimate both individual- and population-level parameters as part of the same estimation procedure. It also allows individual-level parameters to be formulated as random effects so that shrinkage estimators are obtained. In other words, when little information on the parameter value can be gleaned from the data for a patient, the estimate is pulled towards the population average. Thus, for instance, the specification of the overall baseline parasite loads,

Fig. 3. Parasite density profiles (left column) and baseline distributions of parasite stages (right hand column) for selected patients. The left hand column shows both the observed (filled circles) and fitted (open squares) peripheral parasite densities. The right hand column gives the baseline distribution of parasite stages. The vertical arrow beneath the latter indicates the assumed mid-point of sequestration.

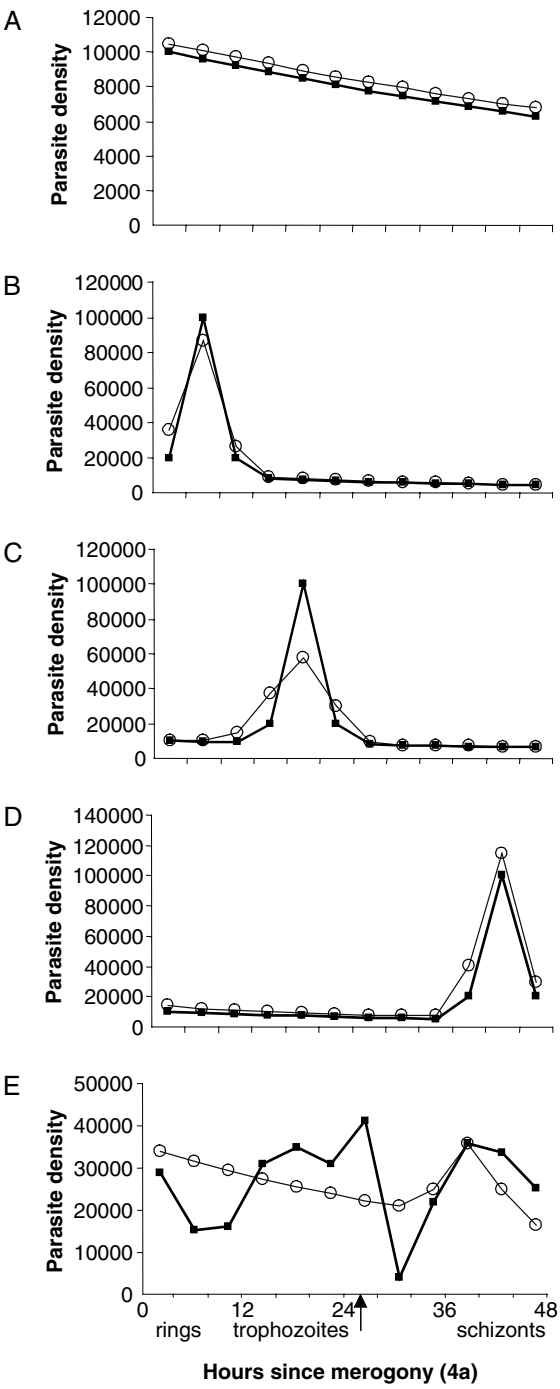


Fig. 6. Estimated baseline status for examples of simulated patients. (A) Asynchronous parasite population. (B) Highly synchronous parasites (predominantly ring-stage). (C) Highly synchronous parasites (predominantly trophozoites). (D) Highly synchronous parasites (predominantly schizonts). (E) Randomly distributed initial parasite population. True (filled squares) and fitted (open circles) parasite densities. The vertical arrow indicates the assumed mid-point of sequestration. Details of the simulated datasets and the estimates of $R_k$ are given in Table 3.

$T_k$, as drawn from a log-normal distribution. Similarly, when the data provide substantial information about $R_k$, our procedure gives a precise estimate of this quantity, when they do not, the posterior distribution of $R_k$ is similar to the prior (a uniform distribution between 1 and 16).

Table 3. Simulated datasets shown in Fig. 6 and corresponding estimates of $R_k$

| | $R_k$ | | Total parasite density at start (/$\mu$l) | |
|---|---|---|---|---|
| | True | Estimate [95% CL] | True | Estimate |
| a. | 2·0 | 2·0 [1·3, 2·9] | 95 984 | 101 834 |
| b. | 16·0 | 15·3 [14·0, 16·0] | 216 371 | 204 014 |
| c. | 2·0 | 2·3 [1·2, 4·0] | 210 664 | 203 645 |
| d. | 16·0 | 12·5 [8·3, 15·8] | 202 763 | 272 310 |
| e. | — | 5·4 [4·8, 6·0] | 320 402 | 328 460 |

The simulation study confirmed that our models do indeed provide reasonable estimates of the patient's status on enrolment, over a wide range of possible initial states. A key quantity that we estimated is $R$, the intrinsic rate of multiplication. This factor has been estimated at 12–15 per cycle in naive hosts during the initial (pre-symptomatic) phase of infection (Fairley, 1947; Cheng *et al.* 1997; Gravenor, McLean & Kwiatkowski, 1995) implying that almost all trophozoites complete schizogony, and most merozoites invade. In contrast White *et al.* (1992) estimated $R$ in ineffectively-treated patients to be 3–10 per cycle, and inferred from the shape of the parasitaemia curves that these lower values for $R$ resulted from reduced merogony or invasion inefficiency, not from clearance of peripheral parasites. Our models suggest that the population of patients included individuals with $R$ values from across the whole range that we allowed. This supports the idea that severe malaria cases are not necessarily characterized by a high intrinsic rate of increase, and that we should look elsewhere for the factors determining severity of *P. falciparum* episodes.

This flexibility represents one substantive advantage over statistical approaches used previously to address this problem. The use of fixed durations of the stages, rather than constant flow-rates represents a further advance over the work of Gravenor *et al.* (1998, 2002). The Gravenor *et al.* (1998) model was formulated by classifying parasites into the two compartments (peripheral and sequestered), and describing the flow rates between the categories with differential equations, corresponding to treating the life-cycle as a Markov process. Its assumption of constant transition rates between the compartments is not equivalent to the usual biologist's model of a constant cycle length (Saul, 1998; Gravenor *et al.* 2002). To address these problems, Gravenor *et al.* (2002) more recently proposed a generalized version of the model that reduces variation in cycle length by increasing the number of compartments, and provides population-level estimates for some

parameters via a second fitting procedure. Maximization of the likelihood for such models can, however, still not always be achieved, and the best fitting parameters imply implausibly high parasite loads for many patients.

If the number of parameters for each patient matches the number of categories the problem is unidentifiable (Gravenor *et al*. 2002), since the model must also include other parameters to allow for pharmacodynamics and parasite multiplication. However, many possible solutions can be rejected as biologically implausible. In the course of developing the present models we were forced, in particular, to reject solutions with very high values of $\kappa$ because they gave total parasite loads for most of the patients in excess of the normal range of erythrocyte counts. Judgement of what are plausible total parasite loads thus introduces constraints that make it possible to solve the problem, but only at the cost of introducing an element of subjectivity into the results.

We also consider it reasonable to require that the parasite density should not vary erratically from one stage to the next (as in the simulated dataset shown in Fig. 6E). This allowed us to exclude solutions with multiple peaks in the distributions of parasite densities across the 12 stage categories, and thus to use only 3 parameters to describe these distributions (together with a further individual-level parameter for the total parasite load). The data still determine the extent of synchronization of the parasites (within the constraint of a fixed value of $\kappa$), selecting between extreme possibilities of highly synchronized parasite populations, uniform distributions corresponding to continuous output of merozoites from the liver, or mixtures of these extremes.

Since the trial was randomized, both total parasite load at baseline, and the proportion sequestered should be similar in patients treated with both drugs. We intended to use comparability of the estimated distribution of stages at baseline as a criterion to evaluate the model. However, for the parameterization that we have reported, while the total load was estimated to be similar the estimated proportion of parasites sequestered was higher in the artemether group. The requirement that each patient should have contributed a total of 8 slides to be included in the analyses may have unintentionally introduced a bias. This requirement meant that only those patients with relatively slow clearance could be included, and difference between the drugs in stage-specificity of their effects means that these form a different subset of the patients at baseline.

The strengths of our model are thus the congruence between the assumptions of the age-stage model and the biology of the parasite, and the fact that it usually provides believable estimates of the total parasite load. However, even with the age-stage model, the sensitivity of the estimates of $R$ and of the estimated stage distributions to small changes in parameterization, caution against uncritical acceptance of the estimates of sequestered loads.

Our primary objective was to develop a model that can be used to cross-validate estimates of the total parasite load made using biochemical assays. For this objective our age-stage model should be adequate, since it seems likely that it gives reasonable estimates of the relative, even if not absolute, numbers of sequestered parasites in each patient, and in particular, that it provides an indication of which patients have a high proportion of sequestered parasites. The results suggest that there is considerable variation between patients in the degree of synchronization, and thus such models should be valuable for studying clinical or patho-physiological outcomes in relation to sequestered load. The model is not useful as a guide for the physician since the whole time-course of parasitaemia measurements must be accrued before the status at baseline can be assessed. The development of biochemical tools to measure the sequestered parasite load thus remains a priority.

This work forms a part of Swiss National Science Foundation project 31-59380.99 (Dynamics of malaria parasites in areas of high transmission). The authors are grateful to two reviewers whose comments enabled us to substantially improve the manuscript.

## REFERENCES

BEST, N., COWLES, M. K. & VINES, K. (1996). CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output. [0.30]. Cambridge, MRC Biostatistics Unit.

CHENG, Q., LAWRENCE, G., REED, C., STOWERS, A., RANFORD-CARTWRIGHT, L., CREASEY, A., CARTER, R. & SAUL, A. (1997). Measurement of *Plasmodium falciparum* growth rates *in vivo*: a test of malaria vaccines. *American Journal of Tropical Medicine and Hygiene* **57**, 495–500.

COLLINS, W. E. & JEFFERY, G. M. (1999). A retrospective examination of sporozoite- and trophozoite-induced infections with *Plasmodium falciparum* in patients previously infected with heterologous species of *Plasmodium*: effect on development of parasitologic and clinical immunity. *American Journal of Tropical Medicine and Hygiene* **61**, 36–43.

EVANS, M., HASTINGS, N. & PEACOCK, B. (2000). ''von Mises Distribution.'' Ch. 41 in *Statistical Distributions*, 3rd Edn. pp. 189–191. Wiley, New York.

FAIRLEY, N. H. (1947). Blood examination and prognosis in acute falciparum malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **40**, 33–48.

GRAVENOR, M. B., MCLEAN, A. R. & KWIATKOWSKI, D. (1995). The regulation of malaria parasitaemia: parameter estimates for a population model. *Parasitology* **110**, 115–122.

GRAVENOR, M. B., LLOYD, A. L., KREMSNER, P. G., MISSINOU, M. A., ENGLISH, M., MARSH, K. & KWIATKOWSKI, D. (2002). A model for estimating total parasite load in falciparum

malaria patients. *Journal of Theoretical Biology* **217**, 137–148.

GRAVENOR, M. B., VAN HENSBROEK, M. B. & KWIATKOWSKI, D. (1998). Estimating sequestered parasite population dynamics in cerebral malaria. *Proceedings of the National Academy of Sciences, USA* **95**, 7620–7624.

HEIDELBERGER, P. & WELCH, P. (1983). Simulation run length control in the presence of an initial transient. *Operations Research* **31**, 1109–1144.

MURPHY, S., ENGLISH, M., WARUIRU, C., MWANGI, I., AMUKOYE, E., CRAWLEY NEWTON, C., WINSTANLEY, P., PESHU, N. & MARSH, K. (1996). An open randomized trial of artemether versus quinine in the treatment of cerebral malaria in African children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**, 298–301.

SAUL, A. (1998). Models for the in-host dynamics of malaria revisited: errors in some basic models lead to large over-estimates of growth rates. *Parasitology* **117**, 405–407.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. & LUNN, D. (2003). WinBUGS. [1.4]. 2003. Cambridge.

TER KUILE, F., WHITE, N. J., HOLLOWAY, P., PASVOL, G. & KRISHNA, S. (1993). *Plasmodium falciparum*: *in vitro* studies of the pharmacodynamic properties of drugs used for the treatment of severe malaria. *Experimental Parasitology* **76**, 85–95.

WARRELL, D. A., MOLYNEUX, M. E. & BEALES, P. F. (eds) (1990). Severe and complicated malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **84** (Suppl. 2), 1–65.

WHITE, N. J., CHAPMAN, D. & WATT, G. (1992). The effects of multiplication and synchronicity on the vascular distribution of parasites in *falciparum* malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **86**, 590–597.

WHITE, N. J. & HO, M. (1992). The pathophysiology of malaria. *Advances in Parasitology* **31**, 83–173.

## APPENDIX: EXPECTED NUMBERS OF PARASITES AT BASELINE

We assume exponential growth in the parasite population, with the survival rate of each age class being constant. Let the total number of stage classes be $A$, the parasite survival during stage class $a$ prior to drug-treatment in host $k$ be $S_{k,a,\mathbf{0}}$, and the number of newly infected erythrocytes arising from one rupturing schizont be $r_k$, so that the net reproductive number, per cycle (of duration $A$ time intervals) is:

$$R_k = r_k \prod_{a=\mathbf{1}}^{A} S_{k,a,\mathbf{0}}. \tag{2.1}$$

For the moment we consider the descendants of a primary population of new rings which arose from merozoites emerging from the liver at rates $x_{\mathbf{1},t_0}, x_{\mathbf{1},t_0+\mathbf{1}} \ldots, x_{\mathbf{1},t_0+A-\mathbf{1}}$ for $A$ time intervals (i.e. exactly 1 cycle) starting at time $t_\mathbf{0}$. We assume these parasites to be sampled at a random time-point, $t: A \leqslant t-t_\mathbf{0} < 2A$. We can use Table A1 (akin to a Lexis diagram) to illustrate the number of parasites in each age class at each time, using the example of $A=4$.

The objective is to determine the expected ratio, $\psi_{a,t} = E[x_{a+\mathbf{1},t}/x_{a,t}]$ for all $a$ and $t$.

First, consider the case where the input rate during the first $A$ intervals is a constant, i.e. $x_{\mathbf{1},t_0} = x_{\mathbf{1},t_0+\mathbf{1}} \ldots = x_{\mathbf{1},t_0+A-\mathbf{1}}$. For each value of $a$, ($a \leqslant A$) there are $A-1$ values of $t$ for which $x_{a+\mathbf{1},t} = S_{k,a,\mathbf{0}} x_{a,t}$ and one value of $t$: $A \leqslant t-t_\mathbf{0} < 2A$ for which $x_{a+\mathbf{1},t} = r_k S_{k,a,\mathbf{0}} x_{a,t}$. The expectation of the ratio over all $t$: $A \leqslant t-t_\mathbf{0} < 2A$ is therefore a weighted mean of these values i.e.:

$$\psi_{a,t} = E\left[\frac{x_{a+\mathbf{1},t}}{x_{a,t}}\right] = S_{k,a,\mathbf{0}}\left(\frac{r_k(A-1)+1}{r_k A}\right), \tag{2.2}$$

and the parasites sampled at any time-point among the $A$ age groups, $a = 1 \ldots A$ are expected to be in the ratio:

$$1 : \psi_{\mathbf{1},t} : \psi_{\mathbf{1},t}\psi_{\mathbf{2},t} : \ldots \prod_{a=\mathbf{1}}^{A-\mathbf{1}} \psi_{a,t}. \tag{2.3}$$

In the conditions applying to our cases on enrolment, we assign $A=12$, and assume $S_{k,a,\mathbf{0}} = 1$ for all $a$ and hence $R_k = r_k$. We define $N_{k,\mathbf{1},\mathbf{0}}$ to be the density of parasites in the first category, and then:

$$N_{k,a,\mathbf{0}} = N_{k,\mathbf{1},\mathbf{0}}\left(\frac{11R_k+1}{12R_k}\right)^{a-\mathbf{1}}. \tag{2.4}$$

We now generalize to $t-t_\mathbf{0} \geqslant 2A$ by noting that the effect of a further cycle of exponential growth is to multiply all the numbers in Table A1 by a factor of $R_k$. This multiplication by a constant has no effect on the ratio $x_{a+\mathbf{1},t}/x_{a,t}$ hence, by induction, we can state that given a constant input rate over one cycle, equation (2.4) applies for any number of completed cycles, i.e. for all $t: t-t_\mathbf{0} \geqslant A$.

Equation (2.2) does not hold for the first cycle after emergence from the liver, hence if the input process continues beyond the first $A$ time-units, parasites undergoing this initial cycle will distort the age distribution of the parasites away from that defined by (2.4). However, as the parasite population expands, newly emerged parasites represent a rapidly decreasing proportion of the total. Mixing of parasites which have undergone different numbers of cycles does not lead to violation of equation (2.4), so the fact that there are different cohorts of parasites contributing to each age group does not invalidate this equation.

In practice we anticipate that the rate of output of merozoites from the liver is not constant, and that small variations in cycle length will occur. Any

Table A1. Expected values of $x_{a,t}$ for the period 4 to 8 time categories after the start of emergence from the liver, assuming 4 stage categories

| | $a=1$ | $a=2$ | $a=3$ | $a=4$ |
|---|---|---|---|---|
| $t=t_0+4$ | $r_k x_{1,t_0}$ | $x_{1,t_0+3}S_{k,1,0}$ | $x_{1,t_0+2}\prod_{\alpha=1}^{2} S_{k,\alpha,0}$ | $x_{1,t_0+1}\prod_{\alpha=1}^{3} S_{k,\alpha,0}$ |
| $t=t_0+5$ | $r_k x_{1,t_0+1}$ | $r_k x_{1,t_0}S_{k,1,0}$ | $x_{1,t_0+3}\prod_{\alpha=1}^{2} S_{k,\alpha,0}$ | $x_{1,t_0+2}\prod_{\alpha=1}^{3} S_{k,\alpha,0}$ |
| $t=t_0+6$ | $r_k x_{1,t_0+2}$ | $r_k x_{1,t_0+1}S_{k,1,0}$ | $r_k x_{1,t_0}\prod_{\alpha=1}^{2} S_{k,\alpha,0}$ | $x_{1,t_0+3}\prod_{\alpha=1}^{3} S_{k,\alpha,0}$ |
| $t=t_0+7$ | $r_k x_{1,t_0+3}$ | $r_k x_{1,t_0+2}S_{k,1,0}$ | $r_k x_{1,t_0+1}\prod_{\alpha=1}^{2} S_{k,\alpha,0}$ | $r_k x_{1,t_0}\prod_{\alpha=1}^{3} S_{k,\alpha,0}$ |

random variation such that, $x_{1,t}=\varepsilon_t x_{1,t_0}$ will result in multiplication of the density of all descendent parasites by the factor $\varepsilon_t$. Thus random variation over time in the input from the liver at time $t$ will result in proportionate random variation in $x_{a,t+\delta t}$ around the ratios expected from (2.4), for all $\delta t$ satisfying $\delta t=Ac+a-1$, where $c$ is an integral number of completed cycles.