American Journal of Epidemiology
Copyright © 2000 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved

Vol. 152, No. 8
Printed in U.S.A.

Bayesian Analysis of an Epidemiologic Model of *Plasmodium falciparum* Malaria Infection in Ndiop, Senegal

Nicole Cancré,[1,2] Adama Tall,[1] Christophe Rogier,[3] Joseph Faye,[1] Ousmane Sarr,[1] Jean-François Trape,[4] André Spiegel,[1] and Frédéric Bois[2,5]

Plasmodium falciparum has a complex transmission cycle. Public health planning and research would benefit from the ability of a calibrated model to predict the epidemiologic characteristics of populations living in areas of malaria endemicity. This paper describes the application of Bayesian calibration to a malaria transmission model using longitudinal data gathered from 176 subjects in Ndiop, Senegal, from July 1, 1993, to July 31, 1994. The model was able to adequately predict *P. falciparum* parasitemia prevalence in the study population. Further insight into the dynamics of malaria in Ndiop was provided. During the dry season, the estimated fraction of nonimmune subjects goes down to 20% and then increases up to 80%. The model-predicted time-weighted average incidences contributed by nonimmune and immune individuals are 0.52 cases per day and 0.47 cases per day, respectively. The median times needed to acquire infection (conversion delay) for nonimmune and immune individuals are estimated at 39 days and 285 days, respectively. *Am J Epidemiol* 2000;152:760–70.

disease transmission; immunity; malaria; models, statistical; *Plasmodium falciparum*

Malaria-induced mortality and morbidity are increasing worldwide (1). New orientations for control of the disease are emphasizing reduction in mortality and morbidity rather than eradication (2, 3). Particular attention is being given today to the development of vaccines. For scientists seeking to isolate potential vaccines, for clinicians evaluating vaccine safety and efficacy, and for public health managers developing malaria control programs, it is important to know the epidemiologic characteristics of the disease and of the populations living in areas of endemicity. For example, in clinical trials organized to test the efficacy of potential vaccines, ways to minimize the proportion of nonsusceptible, eventually immune recruited subjects would improve statistical power.

However, malaria is a complex disease. *Plasmodium falciparum* infection confers only labile and partial immunity. Acquisition of such immunity only reduces the incidence of clinical malaria attacks without preventing infection (4–7); an infected subject may acquire a new infection before recovering from a previous one (a phenomenon known as "superinfection") (8); and immunity is acquired slowly and is a function of exposure to infecting mosquitoes (9). To further complicate the problem, exposure to infectious mosquito bites is difficult to measure (10). Practically, exposure cannot be assessed for each subject, and all individuals are usually considered identically exposed; this constitutes a potential confounding factor for acquisition of immunity. The presence of parasites in individuals gives little information on exposure, because parasitemia can last a long time in the absence of treatment (11). Finally, in some geographic areas, exposure shows marked seasonality and can be very different from one year to the next (4).

An epidemiologic model accounting for these complexities would be a useful tool with which to describe the dynamics of malaria and assess the epidemiologic status of exposed populations (12), answering the needs of both public health planners and malaria researchers. Among the relevant models reported in the literature (8, 12–25), few have been statistically calibrated (i.e., formally fitted to data) because of a lack of extensive longitudinal data and adequate statistical techniques. In two cases (16, 21), three or four of the model parameters were estimated through minimization of χ^2 or G^2 (akin to entropy) criteria of fit. Simulation results from these models have never been presented with confidence intervals allowing an assessment of their reliability. Newly developed Bayesian numerical techniques (e.g., Markov chain Monte Carlo methods) (26–28) offer the possibility of an extensive statistical treatment of complex epidemiologic models, including inference about model parameter values, calculation of confidence intervals for model predictions, model-checking, and hypothesis-

Received for publication August 13, 1998, and accepted for publication December 3, 1999.

Abbreviations: CI, confidence interval; EIR, entomologic inoculation rate.

[1] Institut Pasteur de Dakar, Dakar, Senegal.
[2] Institut National de la Santé et de la Recherche Médicale (INSERM), Unité 444, Faculté de Médecine Saint-Antoine, Université Pierre et Marie Curie, Paris, France.
[3] Institut de Médecine Tropicale–Service de Santé des Armées, Marseille, France.
[4] Institut de Recherche pour le Développement (formerly ORSTOM), Dakar, Senegal.
[5] Institut National de l'Environnement Industriel et des Risques, Verneuil en Halatte, France.
Reprint requests to Dr. Frédéric Bois, INERIS, Parc Alata, BP2, 60550 Verneuil-en-Halatte, France (email: bois@b3e.jussieu.fr).

testing. In this paper, we apply these techniques to the malaria model described by Struchiner et al. (23). This model embodies a series of coherent, plausible research hypotheses about the natural history of *P. falciparum* malaria and offers a synthesis of previous work carried out by MacDonald (8), Dietz et al. (16), and Nedelman (21). We use Markov chain Monte Carlo simulations to fit the model to longitudinal data gathered from the village of Ndiop in a meso-endemic area of Senegal. Model simulations of the underlying infection dynamics are presented.

MATERIALS AND METHODS

Population and data

The data used were gathered in the village of Ndiop (13°41′N, 16°23′W) in the Sahelo-Soudanian region of Senegal. Since 1993, a longitudinal study has been conducted in Ndiop, in which entomologic, parasitologic, and clinical data have been collected. This analysis used data gathered from July 1, 1993, to July 31, 1994. The cohort follow-up was particularly intense from July to October of 1993, when the prevalence of *P. falciparum* parasitemia was highest. The rainy season usually lasts from June to October.

Human data. Only the 176 villagers (of a total of 396) who were continuously present in the village during the study period were included. Exclusions can be considered random with respect to exposure (in particular, it is very unlikely that people left the village because of mosquito bites or to obtain treatment elsewhere), and no bias should have been introduced by the removal of the traveling villagers. On the other hand, traveling villagers had unmeasured exposure to mosquitoes during their outings, and they could have introduced bias if included. A local medical care unit was created to support the study after agreement was reached with the population of the village and the public health authorities. Informed consent was obtained individually from the participants or from their parents (for children); approval was obtained from the Ministère du Plan et de la Coopération and from the Ministère de la Santé Publique. The local medical care unit was provided with basic equipment for malaria diagnosis. A team of four physicians, two technicians, and four fieldworkers stationed in the village was in charge of the contacts with the community and clinical follow-up. They were all trained in the clinical and laboratory diagnosis of malarial infection. Active surveillance consisted of: 1) a daily visit to each villager at home, to record body temperature and clinical symptoms of any type that had occurred during the previous 24 hours; and 2) collection of thick blood smears once per week from July to October of 1993 and once per month from November 1993 to August 1994. Passive surveillance included collection of thick blood smears from subjects reporting to the health unit on any occasion. Special attention was paid to the use of antimalarial drugs in the community, and the population was asked not to use any such drugs without prescription. The only antimalarial drug used was quinine (Quinimax; Sanofi-Labaz, Paris, France), administered at a dose of 25 mg/kg/day for 7 days to the following types of patients: children under 10 years of age with fever (temperature higher than 38°C) and high parasitemia (over 30 trophozoites per 100 leukocytes); pregnant women with clinical symptoms suggestive of malaria; persons with fever and very high parasitemia (over 200 trophozoites per 100 leukocytes); and persons with severe malaria symptoms (coma, etc.).

A total of 5,736 thick blood smears were collected from the 176 villagers studied. To study the natural evolution of parasitemia without interfering with antimalarial treatments, we excluded from the analysis thick blood smears collected within 15 days after the beginning of treatment. As a result, only 5,000 thick blood smears were considered in our analysis. All smears were double-read, once in the field by the technicians and once by an expert microscopist at the Institut de Recherche pour le Développement laboratory in Dakar, whose reading was definitive. Slides were stained using 4 percent Giemsa stain, and up to 200 microscopic oil-immersion fields were examined at a magnification of 100. Asexual-stage parasite (trophozoite) densities were reported as parasite count for 100 leukocytes (detection limit: 0.01 trophozoites for 100 leukocytes). The number, $D(t)$, of trophozoite-positive thick blood smears on any given day for which subjects were seen, given the total number of thick blood smears examined, $M(t)$, constituted a single data point. The entire data set is denoted **D** below. The observed prevalences presented in figure 2 were obtained by dividing $D(t)$ by $M(t)$ at each time t. Table 1 gives the distribution of subjects, positive thick blood smears, and total thick blood smears examined, by age and sex.

Mosquito data. The main anopheline species in Ndiop are *Anopheles arabiensis* and *Anopheles gambiae*. They both contribute to the high endemicity of *P. falciparum* malaria (29). Nighttime captures of mosquitoes attracted to human volunteers were used to sample mosquito populations. Adult mosquitoes were captured on 12 humans for

TABLE 1. Numbers of individuals and thick blood smears included in a study of *Plasmodium falciparum* malaria, by sex and age, Ndiop, Senegal, July 1993–August 1994

Age (years)	No. of subjects in class	No. of trophozoite-positive smears	Total no. of smears examined
		Females	
<1	5	67	160
1–4	9	69	222
5–9	15	207	454
10–14	14	190	396
15–19	8	147	241
20–30	15	171	405
>30	21	214	594
		Males	
<1	6	52	192
1–4	23	265	719
5–9	16	220	485
10–14	8	106	178
15–19	6	106	149
20–30	5	65	168
>30	25	265	637

three consecutive nights each month from July to November of 1993, and then weekly on four humans. Hourly human-bait collections were made on adult volunteers from 7:00 p.m. to 7:00 a.m. The volunteers were always placed at the same locations in the village, half of them indoors and the other half outdoors. The infectivity of captured mosquitoes for *P. falciparum* was assessed by detection of circumsporozoite antigen protein with enzyme-linked immunosorbent assays (30, 31). The number of infected mosquito bites per person (the entomologic inoculation rate (EIR), $h_e(t)$) was hence obtained for each collection day.

Dynamic model

The deterministic compartmental model developed by Struchiner et al. (23) was used to describe the natural history of the infection in humans. A brief summary is given here, since the model has been described elsewhere in full detail (23, 32, 33). The model equations and the definition of each parameter are given in the Appendix.

In the model, the human population is divided into four epidemiologic classes or compartments (figure 1): nonimmune subjects and immune negative subjects, in proportions $X_1(t)$ and $X_3(t)$, respectively; and nonimmune subjects and immune positive subjects, in proportions $Y_2(t)$ and $Y_3(t)$, respectively. These proportions are time-varying, and the model can predict their full time course. Nonimmune positive subjects infectious to mosquitoes, in proportion $Y_1(t)$, are the subset of the nonimmune positives showing sexual-stage (gametocyte) parasitemia. It is assumed that immunity does not totally protect against infection but reduces the probability of becoming infected. Individuals are deemed positive if they show trophozoite parasitemia (as assessed by thick blood smear examination). For our study, the birth and death rate, δ, was equal to zero.

Nonimmune negative individuals receive effective inoculations, in a proportion b_1 of the EIR, and become infected. They show positive parasitemia after an incubation period of N_1 days. An infected person may either acquire immunity or return to the nonimmune negative state. The maximum

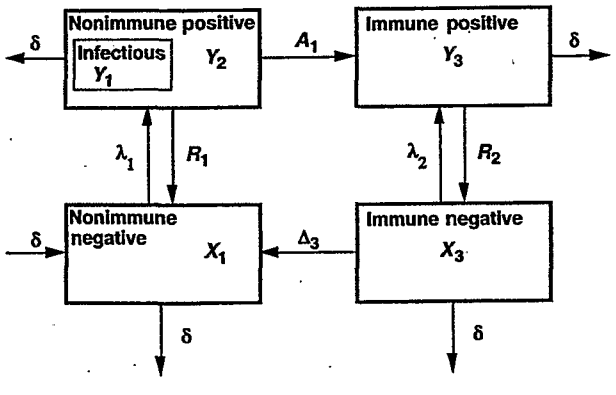


FIGURE 1. Epidemiologic model of *Plasmodium falciparum* malaria in humans (23). The compartments represent the four epidemiologic classes considered in the human population. Arrows indicate transitions among compartments. Symbols are explained in the Appendix (see text).

limiting rate at which immunity can be acquired is α_2. Immune positive individuals are not infectious to mosquitoes and recover more quickly than nonimmune persons. Immune negative individuals receive inoculations, in a proportion b_2 of the EIR resulting in infection. The model specifies that $b_2 < b_1$. Immune negative individuals can lose their immunity if they receive no "booster" inoculation within the time interval τ. Immunity-boosting inoculations are a proportion f of the EIR. Such inoculations prolong an already acquired immunity but do not lead to an established brood of parasites.

An effective inoculation by an infectious mosquito produces one brood of parasites within the human host. Different mosquito bites can each inject a brood into an individual, and those broods are cleared independently of each other. We did not model the mosquito part of the parasite cycle: The EIR defined in Struchiner et al.'s model (23) was replaced by our measurements, linearly interpolated.

Assuming all subjects, at $t = 0$, to be nonimmune and negative ($X_1(0) = 1$) is not realistic for Ndiop. However, neither initial conditions nor the actual "initial time" are known a priori. Preliminary simulations showed that, when starting with realistic conditions, equilibrium is reached in approximately 2.5 years. In that period of time, the system has also practically lost memory of its initial state. Consequently, the initial time was chosen to be December 22, 1990. To acknowledge uncertainty in their values, the initial proportions of individuals in the compartments and initial average numbers of broods in individuals were considered as parameters to estimate. To respect the constraint of the summation to 1 of the initial proportions, we used the reparameterization given in the Appendix (equations 17–19). No data on EIR were available from December 22, 1990, to July 1, 1993, before the beginning of parasitologic monitoring. To supply realistic weekly EIR values for input to the model (equations 9, 10, 14, and 16 in the Appendix) during that period, we used the average of the values recorded in Ndiop from July 1993 to December 1996 for each week.

The model differential equations are nonlinear and include delays. They were integrated numerically, using the "Lsodes" algorithm provided by MCSim software, version 4.2 (34). For given parameter values, integration of the equations shown in the Appendix gives the time courses of its variables ($X_1(t)$, $X_3(t)$, $Y_1(t)$, $Y_2(t)$, $Y_3(t)$, $z_1(t)$, $z_2(t)$, and $z_3(t)$) over any period of time starting on December 22, 1990. The sum, $\tilde{Y}(t)$, of $Y_2(t)$ and $Y_3(t)$ is a model-computed estimate of the instantaneous prevalence of *P. falciparum* parasitemia in humans.

Statistical computations

A Bayesian approach was used to calibrate the model using the counts of *P. falciparum* trophozoite-positive thick blood smears in Ndiop (the data, **D**). Basically, each model parameter (see the list in table 2 and in the Appendix), θ_i, was considered as a random variable and assigned an independent prior distribution, $p(\theta_i)$. Those distributions were updated together to yield a joint posterior distribution, $p(\theta|\mathbf{D})$, such

TABLE 2. Prior distributions adopted for the model parameters in a study of *Plasmodium falciparum* malaria, Ndiop, Senegal, 1993–1994

Parameter	Sources	Distribution	Mean	Standard deviation	Minimum	Maximum
r_1	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.00118*	5†		
r_2	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.0134*	5†		
α_1	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.0108*	5†		
α_2	Molineaux and Gramiccia (9) and Nedelman (21)	Lognormal	0.00026*	5†		
τ	Molineaux and Gramiccia (9), Earle et al. (11), and Anderson and May (12)	Truncated lognormal	365*	2†	90	1,095
N_1	Gilles and Warrell (4) and Molineaux and Gramiccia (9)	Truncated lognormal	15*	2†	5	50
f	—‡	Uniform	0.5	0.289	0	1
b_1	—‡	Uniform	$(1 + b_2)/2$	$(1 - b_2)/\sqrt{12}$	b_2	1
b_2	—‡	Uniform	$f/2$	$f/\sqrt{12}$	0	f
$X_3(0)$	—‡	Uniform	0.5	0.289	0	1
$F_{Y_3(0)}$	—‡	Uniform	0.5	0.289	0	1
$F_{Y_2(0)}$	—‡	Uniform	0.5	0.289	0	1
$z_1(0)$	—‡	Uniform	50	28.9	0	100
$z_2(0)$	—‡	Uniform	50	28.9	0	100
$z_3(0)$	—‡	Uniform	50	28.9	0	100

* Geometric mean.
† Geometric standard deviation (exponential of the standard deviation in log space).
‡ Given the lack of prior information, an uninformative uniform prior was used.

that model-computed time courses of prevalence (using parameter values drawn from that joint posterior) would be compatible with the data. According to Bayes' rule, $p(\boldsymbol{\theta}|\mathbf{D})$ is proportional to the product of the prior distributions by the data likelihood, $p(\mathbf{D}|\boldsymbol{\theta})$, under the model (35, 36).

The prior distributions summarize our knowledge about parameter values before seeing the Ndiop data (table 2). A priori lognormal or truncated lognormal distributions were assigned to several parameters. Their geometric means were set on the basis of the literature (4, 9, 11, 12, 21, 23). Their standard deviations were set by us to large values corresponding to a factor 5 or 2 (the latter for time delays, for which we had better information a priori), with eventual truncation when ranges were suggested by the literature. Uniform distributions over feasible or large regions were assigned in the absence of prior information, particularly for the initial state variables ($X_3(0)$, $z_1(0)$, $z_2(0)$, and $z_3(0)$) or their deconstraining parameters ($F_{Y_2(0)}$ and $F_{Y_3(0)}$).

To define the data likelihood, the observed number, $D(t)$, of trophozoite-positive thick blood smears at time t was assumed to be binomially distributed with parameters $\tilde{Y}(t)$, the model-predicted prevalence ($0 < \tilde{Y}(t) < 1$), and $M(t)$, the total number of thick blood smears counted at t. The joint posterior is therefore of the form

$$p(\boldsymbol{\theta}|\mathbf{D}) \propto \prod_i p(\theta_i) \times \prod_t (\tilde{Y}(t)^{D(t)} \times (1 - \tilde{Y}(t))^{M(t)-D(t)}).$$

Unfortunately, because the dynamic model is nonlinear, there is no known analytical form for $p(\boldsymbol{\theta}|\mathbf{D})$. It is impossible to describe it and report inference about the parameters in a direct way. It remains possible to summarize that distribution by drawing random sets of parameter values using Metropolis sampling (26). This iterative procedure belongs to a class of Markov chain Monte Carlo techniques which has recently received much interest (27, 37–39). Briefly, the algorithm was as follows: At the start of a sampling "chain," all parameters were assigned values sampled from the priors. For any following iteration of the sampler, each component, θ_i, of the parameter vector was eventually updated by drawing a "proposed" new value, θ_i', out of a Gaussian "proposal" distribution centered on θ_i. Values of the joint posterior density at θ_i and θ_i' were then computed using the displayed equation (this required running the differential model to obtain all needed values for $\tilde{Y}(t)$). The two obtained density values were labeled π and π'. If π'/π exceeded 1, the new value θ_i' was accepted and replaced θ_i; otherwise, θ_i' was accepted only with probability π'/π. In the case of rejection of θ_i', the value θ_i was kept. After updating (eventually) all model parameters sequentially (the updating order does not matter in the long run), we recorded their values, therefore completing one iteration of the chain. Iterations were performed until the chain had reached equilibrium, i.e., until all parameters had approximately converged in distribution to $p(\boldsymbol{\theta}|\mathbf{D})$. The standard deviation of the Gaussian proposal distribution was adjusted periodically to yield an acceptance rate of 25 percent (40). The convergence of several Markov chains to $p(\boldsymbol{\theta}|\mathbf{D})$ was assessed using Gelman and Rubin's $\hat{R}$ diagnostic (41). The parameter sets recorded after equilibrium was reached were used to form histograms or compute summary statistics of the posterior distributions for estimands of interest (e.g., marginal parameter distributions, combinations of parameters, or model predictions). Obtaining the posterior distribution of model predictions required running the malaria model once for each parameter set recorded. All of the above computations were performed using MCSim software, version 4.2 (34).

RESULTS

Model fit

Convergence of three independent Markov chain Monte Carlo chains was reached after approximately 40,000 iterations ($\hat{R}$) diagnostic at 1.07 on average, ranging from 1 to 1.2). Fifteen thousand parameter sets were sampled, by keeping one out of every six iterations from an additional 30,000 of each chain (each iteration kept yielded a parameter set). All simulations and inferences presented below were made using this final sample from $p(\theta|\mathbf{D})$.

A good fit to the data was obtained, while maintaining scientifically plausible parameter values. Figure 2 shows the daily trophozoite parasitemia prevalences ($D(t)/M(t)$) together with the corresponding model predictions, $\tilde{Y}(t)$, made with the parameter set having the highest posterior density in the final sample. This model prediction is the "best" of all, but it is also quite representative of the set. Ten other model-predicted time courses of prevalence, obtained using parameter vectors randomly drawn from $p(\theta|\mathbf{D})$, are presented. The differing model trajectories for all of these curves reflect uncertainty in model predictions. However, they all have similar behavior. The posterior 95 percent confidence interval for predictions is also displayed. The rise to the peak (from ~20 percent prevalence to 70–80 percent) is somewhat jagged and is mostly driven by the random biting rate of mosquitoes. The subsequent decrease is smoother and is driven by the gradual recovery of the infected subjects in the dry season. Prevalence returns to approximately 20 percent at the end of the dry season. Running a standard smoothing curve through the data would be purely descriptive and would provide no insight into the underlying dynamics. Our goal was not so much to "fit" the data as to extract from them information about the model parameters.

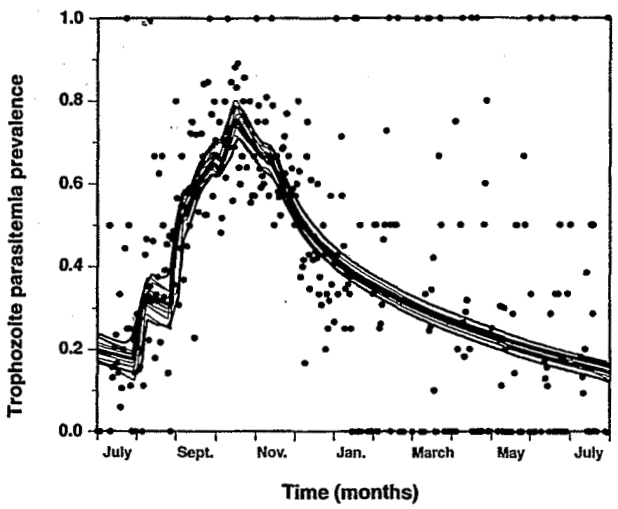


FIGURE 2. Time course of the daily prevalence of *Plasmodium falciparum* malaria (number of thick blood smears with positive trophozoites divided by the total number of smears) among 176 people living in Ndiop, Senegal, from July 1, 1993, to July 31, 1994. The points correspond to observations. The thick line depicts the model-predicted prevalence in the population as a function of time. It was obtained by running the model with the vector of parameter values having the highest posterior density (among a random sample of 15,000 posterior vectors). The thin lines are also model predictions of prevalence, generated with 10 other random parameter vectors drawn from their posterior distribution (see text). The outermost two lines represent the 95% confidence interval for predicted prevalence.

Posterior parameter distributions

The joint posterior distribution of all parameters can be viewed in several dimensions, but for simplicity only the marginal distribution of each parameter is described here. Table 3 summarizes these distributions on the basis of the final parameter sample. For all biologic parameters, the posterior location is noticeably different from the corresponding prior mean. Posterior standard deviations are much lower than specified a priori (compare tables 2 and 3), because important information about those parameters has been extracted from the data.

The median sojourn time ($1/r_1$) of a parasite brood in nonimmune human hosts is 204 days (95 percent confidence interval (CI): 18, 2,000). Although it is lower, this is still compatible with the 850 days previously assumed (9, 21, 23). For immune subjects, this sojourn time is 19 days (30 percent coefficient of variation; 95 percent CI: 11, 30). Immunity appears to affect the life span of the parasite in human hosts.

The recovery rate from infectiousness to mosquitoes among nonimmune positive hosts (α_1) is high but poorly identified. It corresponds to a median half-life of 6 days (95 percent CI: 0.5, 125). The window of infectivity is therefore quite small (as expected from the relative brevity of the prevalence peak during the year).

As indicated by the median value of α_2, 213 days (1/0.0047) are necessary for a human host to acquire immunity to *P. falciparum*. This estimation is quite precise (coefficient of variation of ~10 percent) and is much lower than the a priori value. Simulations indicate that for an inhabitant of Ndiop exposed to seasonal meso-endemic transmission, the numbers of infectious and noninfectious parasite broods present at any time in nonimmune hosts (z_1 and z_2) average approximately 10 (data not shown). Under such conditions (see equation 13 in the Appendix), the actual immunity acquisition delay (A_1) is equal to 214 days on average, which is close to its minimum value. Immunity appears more quickly than expected a priori, but it still takes at least half a year to be in effect.

The interval of time, τ, until an immune host loses immunity in the absence of exposure to infectious mosquito bites is approximately 230 days (95 percent CI: 180, 290). The incubation period, N_1, lasts 22 days, on average (95 percent CI: 15, 28).

Few infectious bites (f = ~3 percent) are able to boost immunity, and this boosting does not seem to be very important a posteriori. The proportion of potentially infectious bites actually resulting in infection is much higher in nonimmune hosts (b_1 = 40 percent; 95 percent CI: 26, 91) than in immune subjects (b_2 = 2.5 percent; 95 percent CI: 1.5, 3.5). Immunity, although it is progressively acquired, seems to efficiently protect against infection.

TABLE 3. Summary of the posterior (fitted) distributions for the model parameters in a study of *Plasmodium falciparum* malaria, Ndiop, Senegal, 1993–1994

Parameter	Median	Mean (SD*)	Geometric mean (GSD*)	2.5th percentile	25th percentile	75th percentile	97.5th percentile
r_1	0.0049	0.030 (0.18)	0.0047 (3.3)	0.0005	0.0026	0.0082	0.057
r_2	0.053	0.055 (0.015)	0.054 (1.3)	0.033	0.045	0.063	0.095
α_1	0.12	0.33 (0.51)	0.12 (4.9)	0.0055	0.04	0.39	2.0
α_2	0.0047	0.0047 (4.6×10^{-4})	0.0047 (1.1)	0.0037	0.0044	0.0051	0.0055
τ	230	235 (30)	230 (1.1)	180	220	260	290
N_1	22	22.5 (2.9)	22 (1.2)	15	21	24	27.5
f	0.030	0.038 (0.051)	0.032 (1.5)	0.019	0.026	0.035	0.076
b_1	0.39	0.43 (0.16)	0.41 (1.4)	0.26	0.33	0.48	0.91
b_2	0.024	0.024 (0.005)	0.023 (1.2)	0.015	0.021	0.027	0.035
$X_3(0)$	0.66	0.62 (0.26)	0.525 (2.0)	0.080	0.42	0.84	0.98
$F_{Y_3(0)}$	0.37	0.42 (0.28)	0.30 (2.7)	0.025	0.18	0.66	0.96
$F_{Y_2(0)}$	0.42	0.45 (0.27)	0.33 (2.6)	0.030	0.22	0.66	0.95
$z_1(0)$	51	50.5 (28)	39 (2.5)	3.7	27	74	96
$z_2(0)$	48	49 (28)	37 (2.5)	3.3	25	72	96
$z_3(0)$	52	51 (28)	38 (2.6)	2.7	27	75	97

* SD, standard deviation; GSD, geometric standard deviation.

The marginal posterior distributions of the sampled initial state variables or their reparameterizations ($X_3(0)$, $F_{Y_2(0)}$, $F_{Y_3(0)}$, $z_1(0)$, $z_2(0)$, and $z_3(0)$) are very close to the corresponding priors. This shows the insensitivity of the model to those parameters: Their values do not appreciably affect the results.

Model predictions of the underlying dynamics

After fitting, the model can be used to simulate various scenarios to better understand the dynamics of *P. falciparum* infection in our study population. Figure 3 shows predictions of the time course of malaria immunity status during the year of our study and the following wet season (July 1993 to December 1994), together with the measured EIR. At the end of the dry season, the population composition is as follows: 63 percent (95 percent CI: 35, 85) nonimmune negative (X_1), 12 percent (95 percent CI: 10, 15) nonimmune positive (Y_2), 1 percent (95 percent CI: 0.7, 2) immune positive (Y_3), and 24 percent (95 percent CI: 5, 50) immune negative (X_3). These proportions vary from year to year as mosquito biting fluctuates in timing and intensity. Nonetheless, some behaviors appear stable. The proportion of nonimmune negative individuals falls very quickly, and practically to zero, as soon as mosquito biting increases. Infected subjects initially transfer to a nonimmune positive status, whose proportion reaches a peak at approximately the same time as the biting rate (with a short delay imposed by incubation). Nonimmune positive subjects then transfer mostly to the immune status. However, the fraction of immune positive individuals does not increase much, and individuals quickly eliminate parasites to become immune negative. Near the end of the dry season, immune negative individuals lose immunity and the fraction of nonimmune negative persons increases quickly. Overall, most subjects move through the four states as the year progresses.

Other epidemiologic data point to the dependence of immunity on continued exposure (9). For example, in a malaria control program consisting of insecticide spraying against mosquitoes and mass antimalarial treatment of the human population for two wet seasons, malaria prevalence during the subsequent wet season was higher than in a control population; the following year, in the absence of intervention, prevalence became similar for the two populations

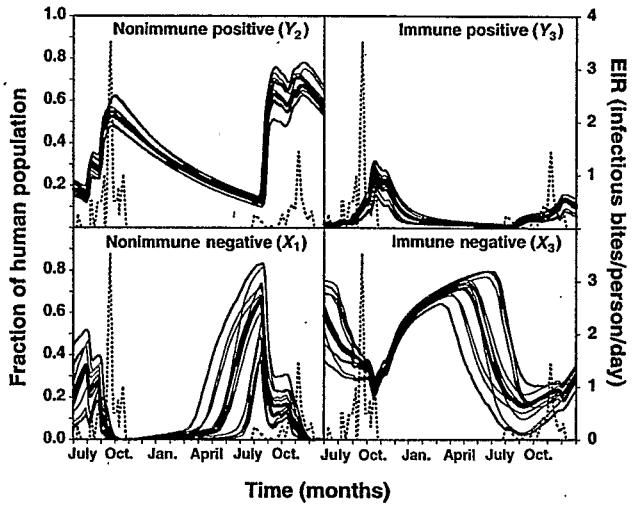


FIGURE 3. Model predictions of the fractions of the human population in four epidemiologic states (see figure 1) with regard to *Plasmodium falciparum* malaria, Ndiop, Senegal, 1993–1994. The thick lines show the predictions obtained by running the model with the vector of parameter values having the highest posterior density. The thin lines show the predictions generated with 10 random posterior parameter vectors. The outermost two lines in each section represent the 95% confidence interval for the predictions. The dotted lines correspond to the entomologic inoculation rate (EIR) for the same period of time.

(9). This behavior, also discussed by Halloran et al. (32), is reproduced by the present model (data not shown).

The incidence of malarial infection is difficult to measure, since it requires identification of new infections in individuals who are potentially already parasitemic. The model can easily provide an estimate for various incidence rates. Figure 4 presents the model-reconstructed instantaneous incidence rates (number of new cases per day) of trophozoite parasitemia contributed by nonimmune and immune individuals during the 1993 wet season in Ndiop (incidence was null during the dry season). These rates correspond to the products $\lambda_1(t) \times X_1 \times 396$ and $\lambda_2(t) \times X_3 \times 396$, respectively (396 being the total size of the population of Ndiop). The time-weighted average incidences, over the year, contributed by nonimmune and immune individuals are approximately the same—0.52 (95 percent CI: 0.40, 0.63) and 0.47 (95 percent CI: 0.30, 0.67) cases per day, respectively. However, the incidence time profiles for these two subpopulations do differ. Early incident cases are contributed mostly by nonimmune individuals and late cases by immune subjects. This is explained by the progressive decline of the nonimmune negative population as the wet season progresses (figure 3). For the whole population, the time-weighted average is estimated at 0.99 (95 percent CI: 0.87, 1.2) cases per day.

Figure 5 gives the immunity acquisition delay and the conversion delays for nonimmune and immune subjects as a function of the EIR. These delays are time-dependent when the EIR varies (see equations 9 and 10 in the Appendix). To avoid this time dependency, the model predictions presented here were computed with constant EIRs. Computations were made with the parameter set having the highest posterior density. When biting is seasonal, as in Ndiop, the curves in figure 5 can still be used to compute approximate delays given yearly average EIRs. The immunity acquisition delay

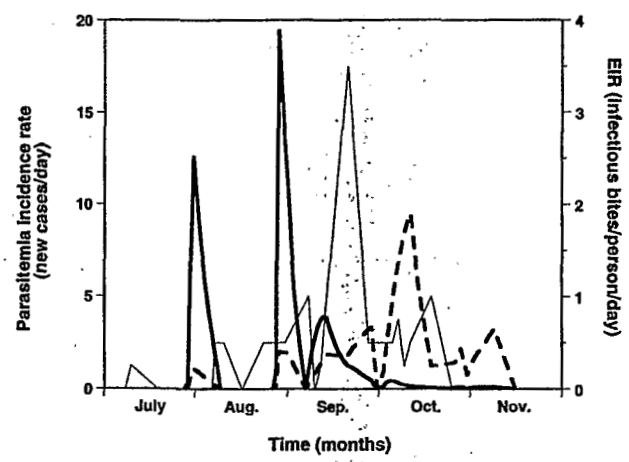


FIGURE 4. Model-reconstructed incidence rate of *Plasmodium falciparum* trophozoite parasitemia in the population of Ndiop, Senegal, during the 1993 wet season. The thick solid line shows the incidence rate contributed by nonimmune subjects; the thick dashed line shows the incidence rate contributed by immune subjects. The thin solid line shows the entomologic inoculation rate (EIR).

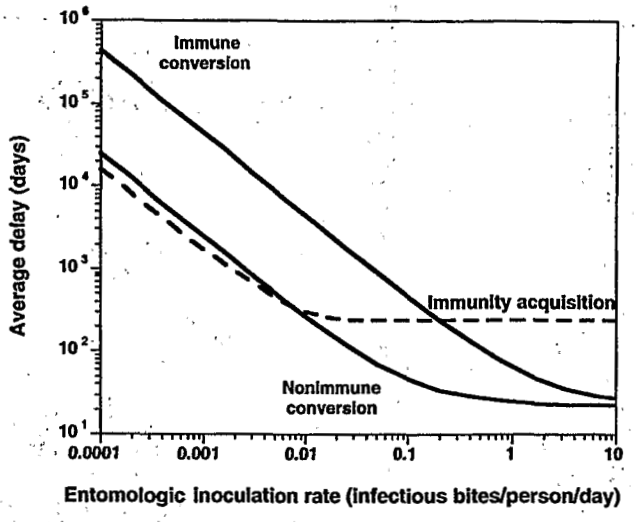


FIGURE 5. Model-predicted immunity acquisition and conversion delays for *Plasmodium falciparum* as a function of the entomologic inoculation rate, Ndiop, Senegal, 1993–1994. Computations were made with the parameter set having the highest posterior density.

($1/A_1$) first decreases proportionally to the inoculation rate. It starts flattening at 0.01 bites per person per day. After that point, it remains at a minimum value of approximately 210 days (equal to $1/\alpha_2$). The conversion delays are the average times it takes a disease-free individual to acquire an infection. They are given by $N_1 + 1/\lambda_1(t)$ and $N_1 + 1/\lambda_2(t)$ for nonimmune and immune subjects, respectively. At low inoculation rates (at least below the Ndiop average of 0.1 potentially infectious bites per person per day), the conversion delay is approximately 16 times (b_1/b_2) lower for immune subjects than for nonimmune subjects. These delays first decrease proportionally to the inoculation rate, and, as it increases, they tend toward a common minimum: the incubation period. For a nonimmune individual in Ndiop in 1993, the median estimated conversion delay was 39 days (95 percent CI: 29, 46). For an immune subject, it was equal to 285 days (95 percent CI: 205, 440). The 95 percent confidence intervals for all delays presented in figure 5 span approximately a factor of 2.

DISCUSSION

This work demonstrates the possibility of statistically calibrating a complex mathematical model with epidemiologic data, using a Bayesian framework. As a result, a reasonable fit of the parasitemia prevalence data was reached, showing that the model is compatible with the observations, and a sample of model parameter values was obtained from their joint posterior distribution. The model was then used to predict quantities that are otherwise difficult to measure, given the current state of knowledge. For example, we obtained estimates of instantaneous and average incidence rates of *P. falciparum* parasitemia. Among the statistical methods available to us, Bayesian updating is particularly appropri-

ate for integrating two forms of information (28, 42, 43): "prior knowledge" from the scientific literature and "data" from field studies. Still, several issues can be raised regarding the data used, the structure of the chosen model, and various assumptions we made.

The data were obtained through intensive follow-up. A potential bias in subject recruitment arose as individuals were offered the opportunity to present themselves for clinical examination. *P. falciparum*-infected individuals with presenting symptoms may have been overrepresented. However, blood samples were analyzed at all self-motivated visits, regardless of whether the visits were related to malaria, and only 16 percent of the samples analyzed were obtained during such visits. We also verified, through examination of residuals after model calibration, that the prevalence of *P. falciparum* parasitemia in self-motivated consultations was not higher than that in systematic screenings. The impact of a potential bias in self-motivated consultations should therefore have been small or nonexistent. We only analyzed data on the prevalence of *P. falciparum* trophozoite parasitemia, but it would be interesting to extend the model to also consider data on the number of clinical malaria attacks.

The model developed by Struchiner et al. (23) offers a reasonable, albeit simplified, description of malaria physiopathology. We did not include the original description by Struchiner et al. of the parasite cycle in mosquitoes. That was not needed, since our data included EIR throughout the year. However, that rate was assumed to be precisely measured and identical for all subjects. This assumption was needed because the full treatment of "error in variables" problems is difficult in the context of large and computationally intensive models. Another important set of modeling assumptions concerns immunity. In the model, an infected person does not necessarily acquire immunity after one inoculation, and immunity can be lost with time. Although the hypothesis of a definitive acquisition of immunity to the different antigenic strains of parasite seen during an individual's lifetime (18) is not explicitly considered, the model does assume that the acquisition of immunity is a function of the number of coinfecting strains.

Our analysis was performed by pooling data on all subjects (but still preserving the longitudinal aspect of the data at the population level). This could be improved by taking into account the age structure of the population—for example, through a hierarchical statistical model (44). This could shed light on age-related differences in susceptibility to *P. falciparum* infection. At this occasion, it might be possible to take into account the fact that the feeding behavior of mosquitoes is affected by a number of host- or environment-related factors (45–48).

According to the model, under conditions similar to those in Ndiop, the fraction of susceptible subjects is the highest at the very end of the dry season, when mosquitoes start biting again. This makes sense given what is known of the natural history of malaria. The advantage of using a calibrated model, assuming it is correct or sufficiently robust, is that it offers a quantitative estimate of this fraction and of the associated uncertainty. Use of such information in vaccination trial design can help researchers assess and improve statistical power. Power calculations show that the effective size of a trial is proportional to the fraction of susceptible subjects (e.g., a study with 10,000 person-days and 50 percent susceptible subjects in each group has the same power as a study with 5,000 person-days and 100 percent susceptible subjects). Location-specific EIRs could be used for input to the model to assess the best time of the year for a study in areas other than Ndiop.

A dynamic perspective on malaria, as embodied in an epidemiologic model able to disentangle time-varying exposures, superinfections, and complex immunity acquisition processes, is essential for a proper analysis of malaria field study data. Too many pitfalls of confounding and bias, difficult to avoid, await standard data analyses. The model analyzed here is by no means complete or perfect, but it offers a reasonable basis for extension and improvement. Several research teams worldwide are currently attempting to improve malaria models. These efforts would benefit from the statistical techniques presented here. Calibrated models can be powerful predictive tools for experimental design and exploration of public health measures.

ACKNOWLEDGMENTS

This work was supported by a grant from the Ministère de la Coopération et du Développement (Paris, France).

The authors acknowledge the administrative and technical support of the Institut Pasteur and Institut de Recherche pour le Développement in Dakar, Senegal, particularly the help of E.-H. Ba. The authors are grateful to the villagers of Ndiop and to the local study team for their collaboration. They also thank Drs. K. Dietz, J. Eisenberg, D. Fontenille, O. Garraud, S. Gupta, and L. Lochouarn for their helpful comments.

REFERENCES

1. Activités antipaludiques: les 40 dernières années. World Health Stat Q 1988;41:64–73.
2. World Health Organization. WHO Expert Committee on Malaria: eighteenth report. Geneva, Switzerland: World Health Organization, 1986.
3. Najera JA. Malaria control: present situation and need for historical research. Parassitologia 1990;32:215–29.
4. Gilles HM, Warrell DA. Bruce-Chwatt's essential malariology. London, United Kingdom: Edward Arnold, 1993.
5. Christophers SR. The mechanism of immunity against malaria in communities living under hyper-endemic conditions. Ind J Med Res 1924;12:273–94.
6. Miller M. Observations on the natural history of malaria in the semi-resistant West African. Trans R Soc Trop Med Hyg 1958; 52:152–67.
7. Garnham PC. Malaria immunity in Africans: effects in infancy and early childhood. Ann Trop Med Parasitol 1949;43:47–61.
8. MacDonald G. The analysis of infection rates in diseases in which superinfection occurs. Trop Dis Bull 1950;47:907–14.
9. Molineaux L, Gramiccia G. The Garki Project. Geneva, Switzerland: World Health Organization, 1980.

10. Smith T, Charlwood JD, Takken W, et al. Mapping the densities of malaria vectors within a single village. Acta Trop 1995; 59:1–18.
11. Earle WC, Perez M, Del Rio J, et al. Observations on the course of naturally acquired malaria in Puerto Rico. Puerto Rico J Public Health Trop Med 1939;14:391–406.
12. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. New York, NY: Oxford University Press, 1991.
13. Bekessy A, Molineaux L, Storey J. Estimation of incidence and recovery rates of *Plasmodium falciparum* parasitaemia from longitudinal data. Bull World Health Organ 1976;54: 685–93.
14. Aron JL. Malaria epidemiology and detectability. Trans R Soc Trop Med Hyg 1982;76:595–601.
15. Aron JL. Mathematical modeling of immunity to malaria. Math Biosci 1988;90:385–96.
16. Dietz K, Molineaux L, Thomas A. A malaria model tested in African savannah. Bull World Health Organ 1974;50:347–57.
17. Bailey NT. The biomathematics of malaria. London, United Kingdom: Griffin and Company Ltd, 1982.
18. Gupta S, Trenholme K, Anderson RM, et al. Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. Science 1994;263:961–3.
19. Milligan PJ, Downham DY. Models of superinfection and acquired immunity to multiple parasite strains. J Appl Probabil 1996;33:915–32.
20. Dutertre J. Etude d'un modèle épidémiologique appliqué au paludisme. Ann Soc Belge Méd Trop 1976;56:127–41.
21. Nedelman J. Inoculation and recovery rates in the malaria model of Dietz, Molineaux, and Thomas. Math Biosci 1984; 69:209–33.
22. Nasell I. On superinfection in malaria. IMA J Math Appl Med Biol 1986;3:211–27.
23. Struchiner CJ, Halloran ME, Spielman A. Modeling malaria vaccines I: new uses for old ideas. Math Biosci 1989;94: 87–113.
24. Dietz K. Mathematical models for transmission and control of malaria. In: Wernsdorfer WH, McGregor I, eds. Principles and practice of malariology. London, United Kingdom: Churchill Livingstone, 1988:1091–133.
25. De Zoysa AP, Mendis C, Gamage-Mendis AC, et al. A mathematical model for *Plasmodium vivax* malaria transmission: estimation of the impact of transmission-blocking immunity in an endemic area. Bull World Health Organ 1991;69:725–34.
26. Smith AF. Bayesian computational methods. Philos Trans R Soc Lond A 1991;337:369–86.
27. Gelman A. Iterative and non-iterative simulation algorithms. Comput Sci Stat 1992;24:433–8.
28. Gelman A, Rubin DB. Markov chain Monte Carlo methods in biostatistics. Stat Methods Med Res 1996;5:339–55.
29. Fontenille D, Lochouarn L, Diatta M, et al. Four years' entomological study of the transmission of seasonal malaria in Senegal and the bionomics of *Anopheles gambiae* and *A. arabiensis*. Trans R Soc Trop Med Hyg 1997;91:647–52.
30. Wirtz RA, Zavala F, Charoenvit Y, et al. Comparative testing of monoclonal antibodies against *Plasmodium falciparum* sporozoites for ELISA development. Bull World Health Organ 1987;65:39–45.
31. Beier JC, Perkins PV, Koros JK, et al. Malaria sporozoite detection by dissection and ELISA to assess infectivity to Afrotropical *Anopheles*. J Med Entomol 1990;27:377–84.
32. Halloran ME, Struchiner CJ, Spielman A. Modeling malaria vaccines II: population effects of stage-specific malaria vaccines dependent on natural boosting. Math Biosci 1989;94: 115–49.
33. Halloran ME, Struchiner CJ. Modeling transmission dynamics of stage-specific malaria vaccines. Parasitol Today 1992;8: 77–85.
34. Bois FY, Maszle D. MCSim: a simulation program. J Stat Software 1997;2. http://www.stat.ucla.edu/journals/jss/v02/i09 (also available at ftp://sparky.berkeley.edu/pub/mcsim).
35. Bernardo JM, Smith AF. Bayesian theory. New York, NY: John Wiley and Sons, Inc, 1994.
36. Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. London, United Kingdom: Chapman and Hall Ltd, 1995.
37. Gelfand AE, Smith AF. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85:398–409.
38. Gelfand AE, Smith AF, Lee T-M. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. J Am Stat Assoc 1992;87:523–32.
39. Gilks WR, Richardson S, Spiegelhalter DJ. Markov chain Monte Carlo in practice. London, United Kingdom: Chapman and Hall Ltd, 1996.
40. Roberts GO, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk metropolis algorithms. Ann Appl Probabil 1997;7:110–20.
41. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7:457–511.
42. Best NG, Spiegelhalter DJ, Thomas A, et al. Bayesian analysis of realistically complex models. J R Stat Soc A 1996;159: 323–42.
43. Givens GH, Hughes JP. A method for determining uncertainty of predictions from deterministic epidemic models. In: Anderson JG, Katzper M, eds. Proceedings of Health Sciences, Physiological and Pharmacological Simulation Studies—1995 Western Multiconference. San Diego, CA: Society for Computer Simulation, 1995:205–10.
44. Lange N, Carlin BP, Gelfand AE. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. J Am Stat Assoc 1992;87:615–32.
45. Glynn J, Bradley D. Inoculum size, incubation period and severity of malaria: analysis of data from malaria therapy records. Parasitology 1995;110:7–19.
46. Lindsay SW, Adiamah JH, Miller JE, et al. Variation in attractiveness of human subjects to malaria mosquitoes (Diptera: Culicidae) in the Gambia. J Med Entomol 1993;30:368–73.
47. Smith T. Proportionality between light trap catches and biting densities of malaria vectors. J Am Mosquito Control Assoc 1995;11:377–8.
48. Tanner M. Kilombero Malaria Project: the level of anti-sporozoite antibodies in a highly endemic malaria area and its relationship with exposure to mosquitoes. Trans R Soc Trop Med Hyg 1992;86:499–504.

APPENDIX

The following equations, from Struchiner et al. (23), describe the transitions among model compartments:

$$\frac{dX_1(t)}{dt} = \delta(t) + R_1(t)Y_2(t) - (\lambda_1(t) + \delta(t))X_1(t) + \Delta_3(t) \quad (1)$$

$$\frac{dX_3(t)}{dt} = R_2(t)Y_3(t) - (\lambda_2(t) + \delta(t))X_3(t) - \Delta_3(t) \quad (2)$$

$$\frac{dY_2(t)}{dt} = \lambda_1(t)X_1(t) - [A_1(t) + R_1(t) + \delta(t)]Y_2(t) \tag{3}$$

$$\frac{dY_3(t)}{dt} = \lambda_2(t)X_3(t) + A_1(t)Y_2(t) - [R_2(t) + \delta(t)]Y_3(t) \tag{4}$$

$$\frac{dz_1(t)}{dt} = \lambda_1(t) - \alpha_1(t)z_1(t) \tag{5}$$

$$\frac{dz_2(t)}{dt} = \alpha_1(t)z_1(t) - r_1z_2(t) \tag{6}$$

$$\frac{dz_3(t)}{dt} = \lambda_2(t) - r_2z_3(t) \tag{7}$$

$$Y_1 = \frac{1 - e^{-z_1}}{1 - e^{-(z_1+z_2)}}Y_2 \tag{8}$$

$$\lambda_1(t) = b_1h_e(t - N_1) \tag{9}$$

$$\lambda_2(t) = b_2h_e(t - N_1) \tag{10}$$

$$R_1 = r_1z_2\frac{e^{-(z_1+z_2)}}{1 - e^{-(z_1+z_2)}} \tag{11}$$

$$R_2 = r_2z_3\frac{e^{-z_3}}{1 - e^{-z_3}} \tag{12}$$

$$A_1 = \alpha_2(1 - e^{-(z_1+z_2)}) \tag{13}$$

$$\Delta_3(t) = \{R_2(t - \tau) \times Y_3(t - \tau) + (f - b_2) \times h_e(t - \tau - N_1) \times X_3(t - \tau)\}e^{-(\bar{h}_b+\delta)\tau}, \tag{14}$$

with:

$$\bar{h}_b = \int_{t-\tau}^{t} \frac{h_b(u)}{\tau}du \tag{15}$$

$$h_b(t) = b_2 \times h_e(t - N_1), \qquad b_2 \leq f \leq 1 \tag{16}$$

$$Y_3(0) = F_{Y_3(0)}(1 - X_3(0)) \tag{17}$$

$$Y_2(0) = F_{Y_2(0)}(1 - X_3(0) - Y_3(0)) \tag{18}$$

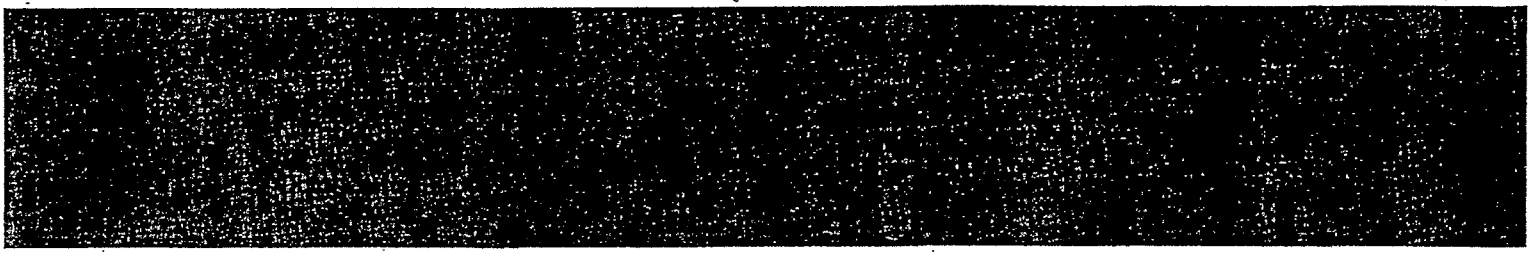
$$X_1(0) = 1 - X_3(0) - Y_3(0) - Y_2(0). \tag{19}$$

The symbols used in the above equations and in the text are defined below (in alphabetical order).

A_1: rate at which immunity to *Plasmodium falciparum* infection is acquired by a human host.
b_1: proportion of bites by infectious mosquitoes on negative nonimmune hosts actually resulting in infection.
b_2: proportion of bites by infectious mosquitoes on negative immune hosts actually resulting in infection.
f: boosting factor (i.e., proportion of bites by infectious mosquitoes on immune hosts resulting in boosted immunity).
$F_{Y_2(0)}$: deconstraining parameter for $Y_2(0)$.
$F_{Y_3(0)}$: deconstraining parameter for $Y_3(0)$.
$h_e(.)$: entomologic inoculation rate (EIR): number of *P. falciparum* infectious bites per human per day.
N_1: *P. falciparum* parasitemia incubation period in humans (in days).
r_1: rate constant of elimination of a brood of parasites by nonimmune positive hosts (in days^{-1}).
r_2: rate constant of elimination of a brood of parasites by immune positive hosts (in days^{-1}).
R_1: recovery rate for nonimmune positive individuals (in days^{-1}).
R_2: recovery rate for immune positive individuals (in days^{-1}).
X_1: proportion of nonimmune negative (i.e., naive) individuals in the population.
$X_1(0)$: initial value (at time 0, i.e., December 22, 1990) of X_1.
X_3: proportion of immune negative individuals.
$X_3(0)$: initial value of X_3.
Y_1: proportion of nonimmune positive individuals potentially infectious for mosquitoes.
Y_2: proportion of nonimmune positive individuals.
$Y_2(0)$: initial value of Y_2.
Y_3: proportion of immune positive individuals.
$Y_3(0)$: initial value of Y_3.
z_1: average number of infectious broods of the parasite per nonimmune positive human host.
$z_1(0)$: initial value of z_1.
z_2: average number of noninfectious broods of the parasite per nonimmune positive human host.
$z_2(0)$: initial value of z_2.
z_3: average number of noninfectious broods of the parasite per immune positive human host.
$z_3(0)$: initial value of z_3.
α_1: recovery rate from infectiousness to mosquitoes among nonimmune positive hosts (in days^{-1}).
α_2: maximum rate at which immunity to *P. falciparum* infection can be acquired by a human host (in days^{-1}).
δ: death and birth rate in the human population (in days^{-1}).
Δ_3: daily fraction of immune negative subjects losing immunity.
λ_1: infection rate for nonimmune negative subjects (probability per day of such a subject's becoming infected).
λ_2: infection rate for immune negative subjects (probability per day of such a subject's becoming infected).
τ: time delay needed for an immune host to lose immunity in the absence of exposure to infection (in days).

ISSN 0002-9262
Printed in the U.S.A.

American Journal of EPIDEMIOLOGY



ORIGINAL CONTRIBUTIONS

Continued on Inside Front Cover