

A stochastic model of malaria transition rates from longitudinal data: considering the risk of “lost to follow-up”

B L VERMA, S K RAY, AND R N SRIVASTAVA

From the Departments of Social and Preventive Medicine, Maharani Laxmi Bai Medical College, Jhansi, and Statistics, Lucknow University, Lucknow, India

ABSTRACT A model, using stochastic processes, is developed to estimate some epidemiological parameters of malaria in a homogeneous population from longitudinal data. Assessments of transition probabilities from one state of health to the other are made taking “lost to follow-up” as a competing risk. The model is based on the assumptions that individuals are transferred at constant rate between states, and only one transition is possible between two consecutive surveys. It shows a good fit to the observed data; the model is simple to understand and can easily be used if computer facilities are not available.

In developing countries where literacy rates are low the task of conducting longitudinal studies on malaria involving blood collection for examination of the total population is notoriously difficult. In such populations, despite the efforts of the survey team, satisfactory coverage, of the sampled population is often not achieved and in subsequent follow-ups a considerable proportion of the population studied initially remains uncontacted. It happens for many reasons, such as “individuals not available at the time of the visit,” “available but refused to co-operate,” “moved to some other place permanently,” “could not be traced,” or “dead.” Such individuals form a category of “lost to follow-up.”

Individuals who are negative for malaria initially may become positive and those found positive in the first survey may, similarly, be negative in the second survey. Furthermore, each individual, irrespective of his initial state, is at risk of being lost to follow-up.

In the present paper an attempt has been made to formulate a simple stochastic model for studying malaria transition probabilities and estimating daily malaria parasite incidence and recovery rates from follow-up data while considering lost to follow-up as a competing risk. Application of the model and numerical results have also been illustrated, based on data collected in a longitudinal study in a small rural population of Uttar Pradesh.

The model

Let a homogeneous population be studied for malaria longitudinally. Three states of health, between two surveys, may be postulated for each individual—namely, (a) S_1 : malaria negative state, (b) S_2 : malaria positive state, and (c) S_3 : lost to follow-up state including death from causes not connected with malaria. The time point of initial survey can be considered to be zero and of subsequent follow-up as t ; the length of $(0, t)$ would thus be t . The transition probabilities during t will depend on intensities of risk; these have been taken to be constant as given below.

State S_i	State S_j		
	S_1	S_2	S_3
S_1	–A	a_1	a_2
S_2	b_1	–B	b_2

(1)

Where $(a_1 + a_2) = A$; and $(b_1 + b_2) = B$.

ASSUMPTIONS

The formulation of the model considers the following assumptions.

(1) The entire study population is exposed to the risk of malaria and risk of lost to follow-up is also present at the time of follow-up.

(2) Individuals are transferred from one state to the other at constant rates and their present transitions are not affected by past ones.

(3) Probability of making one transition in a unit of time is very small and that of making two transitions or more is negligible.

Let $p_{ij}(t)$ denote transition probabilities between the states, for $i = 1, 2$ and $j = 1, 2, 3$ in the interval $(0, t)$ with initial conditions $p_{ii}(0) = 1$ and $p_{ij}(0) = 0$ for $i \neq j$. To define $p_{ij}(t)$ in terms of risk parameters a and b the following differential equations can be produced.¹

$$\begin{aligned}
 p'_{11}(t) &= -A p_{11}(t) \text{ and } p'_{22}(t) = -B p_{22}(t) \\
 p'_{12}(t) &= a_1 p_{11}(t) - B p_{12}(t); p'_{21}(t) = b_1 p_{22}(t) - A p_{21}(t) \\
 p'_{13}(t) &= a_2 p_{11}(t) + b_2 p_{12}(t); \\
 p'_{23}(t) &= b_2 p_{22}(t) + a_2 p_{21}(t)
 \end{aligned}
 \tag{2}$$

Solution² of these differential equations gives

$$\begin{aligned}
 p_{11}(t) &= e^{-At} \\
 p_{12}(t) &= \frac{a_1}{(B - A)} \left[e^{-At} - e^{-Bt} \right] \\
 p_{13}(t) &= \frac{a_2}{A} \left[1 - e^{-At} \right] + \frac{a_1 b_2}{AB} \left[1 - \frac{(Be^{-At} - Ae^{-Bt})}{(B - A)} \right] \\
 \text{and, similarly,} \\
 p_{22}(t) &= e^{-Bt} \\
 p_{21}(t) &= \frac{b_1}{(B - A)} \left[e^{-At} - e^{-Bt} \right] \\
 p_{23}(t) &= \frac{b_2}{B} \left[1 - e^{-Bt} \right] + \frac{b_1 a_2}{AB} \left[1 - \frac{(Be^{-At} - Ae^{-Bt})}{(B - A)} \right]
 \end{aligned}
 \tag{3}$$

ESTIMATION OF PARAMETERS

Consider a homogeneous population of size N . If N_i and N_{ij} are the numbers representing individuals entering into state S_i and total transitions from S_i to S_j respectively in time t , for $i = 1, 2$ and $j = 1, 2, 3$ the likelihood of the observed path can be written as

$$L = C \prod_i N_i \prod_j \left[p_{ij}(t) \right]^{N_{ij}}
 \tag{4}$$

Where C is constant. Solving (4), we get maximum likelihood estimates of

$$p_{ij}(t) = \frac{N_{ij}}{N_i} = f_{ij} \text{ (say)}
 \tag{5}$$

Solution of equations (3) and (5) leads to the following estimates of transition rates.

$$\begin{aligned}
 \hat{a}_1 &= \frac{f_{12}}{t(f_{11} - f_{22})} (\log_e f_{11} - \log_e f_{22}) \\
 \hat{b}_1 &= \frac{f_{21}}{t(f_{11} - f_{22})} (\log_e f_{11} - \log_e f_{22}) \\
 \hat{a}_2 &= \frac{\log_e f_{11}}{t} - \frac{f_{12}}{t(f_{11} - f_{22})} (\log_e f_{11} - \log_e f_{22}) \\
 \hat{b}_2 &= \frac{\log_e f_{22}}{t} - \frac{f_{21}}{t(f_{11} - f_{22})} (\log_e f_{11} - \log_e f_{22})
 \end{aligned}
 \tag{6}$$

Here, \hat{a}_1 and \hat{b}_1 may be interpreted as the estimates of daily "gross" malaria parasite incidence and recovery rates respectively; similarly, equation (3) would be gross or crude transition probabilities. The adjectives gross or crude emphasise that the contemplated transfers from S_i to S_j are affected by other risks competing with this particular risk. Naturally, presence of risks $S_1 \rightarrow S_2 \rightarrow S_3$ and $S_2 \rightarrow S_1 \rightarrow S_3$ diminishes the risk of transfer of individuals from S_1 to S_2 and S_2 to S_1 respectively.

Numerical illustrations

DATA SOURCE

Because the appropriate data to illustrate the applicability of the model were not available in published records, a house-to-house longitudinal study on malaria, consisting of three consecutive surveys at an average interval of four weeks, was undertaken in two villages of Uttar Pradesh having a population of 1830 individuals in 278 families. Methods of survey and other relevant details are given elsewhere (Srivastava RN, *et al*, unpublished data).

Despite our best efforts, only 937 (51.2%)

individuals could be studied for malaria in the initial survey; only individuals covered in the initial survey were followed up the second time, and those covered at the second time were followed up thereafter. Antimalaria drugs, such as chloroquine and camaquine etc were not given to the cases of fever encountered during the first two surveys to enable us to study natural transition-patterns as far as possible; the possibility of taking such drugs by other sources can not be ruled out. No urine test could be performed in view of the limited facilities available. Though surveys were conducted at an average interval of four weeks, analysis showed a mean difference of 30 days between two surveys per person, thus giving a value 30 to the parameter t in the model. Furthermore, individuals studied successfully at two consecutive surveys have been used in the present analysis; all such people for surveys I, II, and III were pooled together to illustrate the applicability of the model.

Results

Table 1 gives observed monthly transition frequencies between states. Estimated values of the risk parameters are shown in table 2; these values, indeed, are the infinitesimal transition probabilities of individuals from one state to the other. Analysis

showed (table 2) that the incidence of overall daily gross malaria and recovery rates were 8 of 10 000 and 26 of 1000 respectively. The gross expected equilibrium³ parasite rate was found to be 3.0% against an observed value of 3.6%. "Expected duration of positive episode" was 38 days.

Monthly gross transition probabilities from S_1 to S_j were also estimated and compared with respective observed ones; these two types of probabilities were almost the same, clearly indicating a good fit of the model to the observed data.

Discussion

The model evolved here considers risk of lost to follow-up competing with the other two defined risks of transitions between states. It attempts to estimate some important epidemiological parameters of malaria—namely, incidence and recovery rates—from longitudinal data in the presence of a familiar problem—incomplete follow-up. Formulation of the model is based on the consideration that parallel to the force of malaria infection there are two other forces—namely, reversible force and force of lost to follow-up simultaneously acting on the population.

The present model assumes probability of two or more transitions between states being negligible.

Table 1 Observed frequencies of transitions between states in 30-day periods (based on surveys I-III)

Sex	No of "negatives" studied	Transitions from state S_1 to S_j (for $j = 1,2,3$)			No of "positives" studied	Transitions from state S_2 to S_j (for $j = 1,2,3$)		
		$f_{11} = \frac{N_{11}}{N_1}$	$f_{12} = \frac{N_{12}}{N_1}$	$f_{13} = \frac{N_{13}}{N_1}$		$f_{22} = \frac{N_{22}}{N_2}$	$f_{21} = \frac{N_{21}}{N_2}$	$f_{23} = \frac{N_{23}}{N_2}$
Men	580	210/580	5/580	365/580	29	5/29	7/29	17/29
Women	648	260/648	4/648	384/648	18	6/18	4/18	8/18
Total	1228	470/1228	9/1228	9/1228	47	11/47	11/47	25/47

Table 2 Estimated values* of risk parameters, observed as well as expected equilibrium parasite rates and expected duration of positive episodes

Sex	Values of risk parameters						Observed parasite rate (%)	Expected [†] equilibrium parasite rate (%)	Expected [‡] duration of positive episodes
	\hat{a}_1	\hat{b}_1	\hat{a}_2	\hat{b}_2	\hat{A}	\hat{B}			
Men	0.0011	0.0315	0.0327	0.0271	0.0339	0.0586	4.53	3.43	31.8
Women	0.0006	0.0202	0.0299	0.0164	0.0304	0.0366	2.79	2.71	49.4
Total	0.0008	0.0258	0.0312	0.0226	0.0320	0.0484	3.62	3.02	38.8

* t has been taken to be 30 days.
[†]Based on surveys I-III.
[‡]Percentage value of $\frac{\hat{a}_1}{\hat{a}_1 + \hat{b}_1}$ gives "expected equilibrium parasite rate."

$\frac{1}{\hat{b}_1}$ provides "expected duration of positive episodes."

Thus the approach considered here is slightly different from that taken into practice usually² where $p_{11}(t)$ and $p_{12}(t)$ in equations (2) contain extra terms, namely $b_1 p_{12}(t)$ and $a_1 p_{21}(t)$ respectively, thus representing a second transition. This leaves an impression that the neglect of these terms in the present model may not matter much over one day but may be so over 30 days—a period that has been considered in the present longitudinal study. To see its effect on $p_{11}(t)$, the form of solution of differential equations representing possibility of further transition was derived; the details have been omitted here for the sake of brevity. Estimation of risk parameters in this case, however, seemed extremely difficult as further solution of necessary equations containing two exponentials resulted in the voluminous calculations. Similar problems were also realised by Fix and Neyman¹ in their model on patients with cancer. Using estimated values of risk parameters a and b from the present model, in equations containing exponentials to derive $p_{11}(30)$, it was seen that discrepancies in probabilities of transitions, derived by two approaches were rather negligible. Furthermore, the maximum likelihood estimates of the risk parameters obtained in the present model are simple to understand. Such estimates using similar criteria have been worked-out by other workers^{1,4,5} too.

Our model shows an appreciably good fit to the observed data. Nevertheless, since the formulation of the model does not consider the possibility of a second transition in time t , it would predict the actual malaria transition rates in cases where the period between two consecutive surveys is not very long. Otherwise, it may show some deviated results. Our results clearly indicate that this model would provide acceptable results for $t \leq 30$.

Results on gross incidence and recovery rates of malaria in the population showed that eight people per 10 000 population acquired malaria infection including relapses, and 26 people per 1000 infected population got rid of the infection a day in the presence of risk to lost to follow-up of cases. Here, since malaria positivity was based on observed parasitaemia, the recovery rates and lapses were derived from the blood slide examination.

Data on incidence and recovery rates of malaria

particularly of the type gathered here are hardly available for any part of our country to allow further interpretation of results. Bekessy *et al.*,⁴ suggested a method for estimating incidence and recovery rates of *Plasmodium falciparum* parasitaemia from follow-up data on the basis of only two states—malaria positive and negative states only. Their method, however, does not take into account the presence of risk of lost to follow-up—a phenomenon that has often been seen in longitudinal studies. Our model deals with a more general case and thus would have a wider applicability to the actual field situations.

The model would have provided more precise results if it were applied to homogeneous data. In fact, homogeneity in the present case was considered for sex and area only, and even the two important variables—age and species—could not be considered in view of the small number of malaria-positive cases detected. Our studied population, therefore, was not strictly homogeneous. Furthermore, data used to illustrate the applicability of the model belonged to a cross-section of a population, small in size, hence the observed results have their own limitations. In studies such as the present one of much greater importance is the sampling variation of the proportions concerned, which clearly indicates the need for population-based studies to be much larger than this.

References

- ¹Fix E, Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. *Hum Biol* 1951, **28**: 205-41.
- ²Chiang CL. *Introduction to stochastic process in biostatistics*. New York: John Wiley and Sons, 1968: 70-110.
- ³Molineaux L, Gramica G. *The Garki project: research on epidemiology and control of malaria in the Sudan Savannah of West Africa*. Geneva: World Health Organisation, 1980.
- ⁴Bekessy A, Molineaux L, Storey J. Estimation of incidence and recovery rates of *Plasmodium falciparum* parasitaemia from longitudinal data. *Bull WHO* 1976; **54**: 685-93.
- ⁵Verma BL, Ray SK, Srivastava RN. Stochastic approach to the estimation of infective force and malaria parasite incidence rate in infants from longitudinal data. *J Commun Dis* 1980; **12**: 118-25.