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.

<table>
<thead>
<tr>
<th>State $S_1$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>$-A$</td>
<td>$a_1$</td>
<td>$a_2$</td>
</tr>
<tr>
<td>$S_2$</td>
<td>$b_1$</td>
<td>$-B$</td>
<td>$b_2$</td>
</tr>
</tbody>
</table>

(1)

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

153
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 \( p_i(t) \) denote transition probabilities between the states, for \( i = 1,2 \) and \( j = 1,2,3 \) in the interval \((0, t)\) with initial conditions \( p_{i0}(0) = 1 \) and \( p_{i0}(0) = 0 \) for \( i \neq j \). To define \( p_{ij}(t) \) in terms of risk parameters \( a \) and \( b \) the following differential equations can be produced.\(^1\)

\[
\begin{align*}
p_{11}(t) &= -A p_{11}(t) \quad \text{and} \quad p_{22}(t) = -B p_{22}(t) \\
p_{12}(t) &= a_1 p_{11}(t) - B p_{22}(t); \quad p_{21}(t) = b_1 p_{22}(t) - A p_{11}(t) \\
p_{13}(t) &= a_2 p_{11}(t) + b_2 p_{22}(t); \\
p_{23}(t) &= b_2 p_{22}(t) + a_2 p_{11}(t)
\end{align*}
\]

Solution\(^3\) of these differential equations gives

\[
\begin{align*}
p_{11}(t) &= e^{-At} \\
p_{12}(t) &= \frac{a_1}{B - A} \left[ e^{-At} - e^{-Bt} \right] \\
p_{21}(t) &= \frac{a_2}{A} \left( 1 - e^{-At} \right) + \frac{a_1 b_2}{AB} \left( 1 - \frac{(Be^{-At} - Ae^{-Bt})}{B - A} \right) \\
p_{22}(t) &= e^{-Bt} \\
p_{23}(t) &= \frac{b_1}{B} \left[ e^{-Bt} - e^{-At} \right] + \frac{b_1 a_2}{AB} \left( 1 - \frac{(Be^{-At} - Ae^{-Bt})}{B - A} \right)
\end{align*}
\]

and, similarly,

\[
\begin{align*}
p_{13}(t) &= e^{-At} \\
p_{12}(t) &= \frac{a_1}{B - A} \left[ e^{-At} - e^{-Bt} \right] \\
p_{21}(t) &= \frac{a_2}{A} \left( 1 - e^{-At} \right) + \frac{a_1 b_2}{AB} \left( 1 - \frac{(Be^{-At} - Ae^{-Bt})}{B - A} \right) \\
p_{23}(t) &= e^{-Bt} \\
p_{23}(t) &= \frac{b_1}{B} \left[ e^{-Bt} - e^{-At} \right] + \frac{b_1 a_2}{AB} \left( 1 - \frac{(Be^{-At} - Ae^{-Bt})}{B - A} \right)
\end{align*}
\]

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 at time \( t \), for \( i = 1,2 \) and \( j = 1,2,3 \) the likelihood of the observed path can be written as

\[
L = C \prod_{i,j} n_{ij} \left[ p_j(t) \right] N_{ij}
\]

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

\[
p_{ij}(t) = \frac{N_{ij}}{N_i} = f_{ij} \quad \text{(say)}
\]

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

\[
\begin{align*}
\hat{a}_1 &= \frac{f_{12}}{t} (\log e_{f_{11}} - \log e_{f_{22}}) \\
\hat{b}_1 &= \frac{f_{21}}{t} (\log e_{f_{11}} - \log e_{f_{22}}) \\
\hat{a}_2 &= \frac{\log e_{f_{11}}}{t} - \frac{f_{12}}{t} (\log e_{f_{11}} - \log e_{f_{22}}) \\
\hat{b}_2 &= \frac{\log e_{f_{21}}}{t} - \frac{f_{21}}{t} (\log e_{f_{11}} - \log e_{f_{22}})
\end{align*}
\]

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 \( S_2 \rightarrow S_1 \rightarrow S_2 \) 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, \textit{et al}, unpublished data).

Despite our best efforts, only 937 (51.2%)
A stochastic model of malaria transition rates from longitudinal data

Based on Percentage

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 1,000 respectively. The gross expected equilibrium 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 to S 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)

<table>
<thead>
<tr>
<th>Sex</th>
<th>No of &quot;negatives&quot; studied</th>
<th>Transitions from state S, to S, (for j = 1, 2, 3)</th>
<th>Transitions from state S, to S, (for j = 1, 2, 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( f_{1} = \frac{N_{1}}{N_{j}} )</td>
<td>( f_{a} = \frac{N_{a}}{N_{j}} )</td>
</tr>
<tr>
<td>Men</td>
<td>580</td>
<td>210/580</td>
<td>5/580</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/580</td>
<td>365/580</td>
</tr>
<tr>
<td>Women</td>
<td>648</td>
<td>260/648</td>
<td>4/648</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4/648</td>
<td>384/648</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/648</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>1228</td>
<td>470/1228</td>
<td>9/1228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/1228</td>
<td>47</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Sex</th>
<th>Values of risk parameters</th>
<th>Observed parasite rate (%)</th>
<th>Expected equilibrium parasite rate (%)</th>
<th>Expected duration of positive episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\alpha}<em>{1}, \hat{\beta}</em>{1}, \hat{\alpha}<em>{2}, \hat{\beta}</em>{2}, \lambda, \beta )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>0.0011</td>
<td>0.0315</td>
<td>0.0327</td>
<td>0.0271</td>
</tr>
<tr>
<td>Women</td>
<td>0.0006</td>
<td>0.0320</td>
<td>0.0299</td>
<td>0.0164</td>
</tr>
<tr>
<td>Total</td>
<td>0.0008</td>
<td>0.0258</td>
<td>0.0312</td>
<td>0.0226</td>
</tr>
</tbody>
</table>

* has been taken to be 30 days.
† Based on surveys I–III.
‡ Percentage value of \( \hat{\alpha}_{1} \) gives "expected equilibrium parasite rate."
Thus the approach considered here is slightly different from that taken into practice usually\(^4\) where \(p_{11}(t)\) and \(p_{22}(t)\) in equations (2) contain extra terms, namely \(b_1 p_{11}(t)\) and \(a_1 p_{11}(t)\) 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 Neyman\(^1\) 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