A mathematical model of seropositivity to malaria antigen, allowing seropositivity to be prolonged by exposure

Samuel Bosomprah

**Abstract**

**Background:** Malaria transmission intensity is traditionally estimated from entomological studies as the entomological inoculation rate (EIR), but this is labour intensive and also raises sampling issues due to the large variation from house to house. Incidence of malaria in the control group of a trial or in a cohort study can be used but is difficult to interpret and to compare between different places and between age groups because of differences in levels of acquired immunity. The reversible catalytic model has been developed to estimate malaria transmission intensity using age-stratified serological data. However, the limitation of this model is that it does not allow for persons to have their seropositivity boosted by exposure while they are already seropositive. The aim of this paper is to develop superinfection mathematical models that allow for antibody response to be boosted by exposure.

**Method:** The superinfection models were fitted to age-stratified serological data using maximum likelihood method.

**Results:** The results showed that estimates of seroconversion rate were higher using the superinfection model than catalytic model. This difference was milder when the level of transmission was lower. This suggests that the catalytic model is underestimating the transmission intensity by up to 31%. The duration of seropositivity is shorter with superinfection model, but still seems too long.

**Conclusion:** The model is important because it can produce more realistic estimates of the duration of seropositivity. This is analogous to Dietz model, which allowed for superinfection and produced more realistic estimates of the duration of infection as compared to the original Ross-MacDonald malaria model, which also ignores superinfection.

**Background**

Measurement of malaria transmission intensity is important for several reasons. It can be used to assess the impact of public health measures on transmission and to understand epidemiological patterns, such as the age distribution of malaria illness. It can also be used for planning intervention studies and interpreting their results. Malaria transmission intensity is traditionally estimated from entomological studies as the entomological inoculation rate (EIR) but this is labour intensive and also raises sampling issues due to the large variation from house to house. Incidence of malaria in the control group of a trial or in a cohort study can be used but is difficult to interpret and to compare between different places and between age groups because of differences in levels of acquired immunity. For common viral infections such as measles, persons can be classed simply as either susceptible or immune, and the force of infection estimated from incidence data, allowing for the fact that individuals become immune after an infection, an approach first developed by [1] who introduce the catalytic model. For common viral diseases such as measles the presence of antibodies indicates the person has been infected and is immune to subsequent infection, the force of infection can then be estimated directly from antibody prevalence by age in cross-sectional serological surveys. This approach has been used to estimate the force of infection for measles and other infections [2-4], using an extension of Muench’s method, which allows that the force of infection may vary with age, but assuming that immunity is lifelong.

Drakeley et al. [5] have developed the use of serological data for estimating malaria transmission intensity, using the concentration of antibodies to MSP119 measured in
surveys of all age groups. They employed a reversible catalytic model, which assumes a constant rate of seroconversion (SCR) and assumes a constant rate of reversion to the seronegative state, independent of the level of transmission. They fitted this model to age-stratified serological data from 12 sites in Tanzania with varying entomological inoculation rate. The model was constrained to fit a single value for the annual rate of reversion to the seronegative state, which was estimated as 0.01393 per year, suggesting that antibodies persist for an average time of 72 years. The authors discussed possible reasons for this long duration. One possible further contributory factor for the estimated long duration is that the model does not incorporate boosting. In high transmission areas, exposure while seropositive prolongs the seropositive state. This feature can be incorporated, allowing each exposure to infection to prolong the seropositive state using a model of superinfection. The aim of this paper is to develop a mathematical model of seropositivity to malaria antigens that allows seropositivity to be prolonged by exposure. The model was validated against the field data used in Drakeley et al. [5]. It was also validated against field data from Bioko, where a change in transmission has been established [6].

Methods
Ethical statement
Ethical approval was obtained from the institutional review boards of the National Institute of Medical Research of Tanzania, Kilimanjaro Christian Medical Centre, and the London School of Hygiene and Tropical Medicine [5].

Data source
Two datasets have been used for assessing the malaria transmission intensity. The study design and methods of data collection are similar. The first dataset involved serum samples collected on 250 people in each of 12 villages in three transects (North Pare, South Pare, and West Usambara) with different transmission intensities in Tanzania. The study design has been described in detail elsewhere [5]. The second dataset is a subset of serum samples collected on 250 people in each of 12 villages in three transects (North Pare, South Pare, and West Usambara) with different transmission intensities. These equations can be solved by a standard method, using generating functions [11].

The basic superinfection model with seroconversion rate, \( \lambda \), assumed constant over time has been derived as:

\[
P(a) = 1 - \exp\left( -\frac{\lambda}{r} (1 - \exp(-ra)) \right)
\]

Statistical analysis
The superinfection models:
The concept of superinfection has been used to describe periods of infection prolonged by repeated exposure to infection [7-10]. Persons in endemic areas often have pre-existing partial immunity. But when these persons are removed from exposure the immunity can be lost gradually. When the person is re-exposed while seropositive the level of antibody response can be boosted. A simple way to allow for the antibody response to be boosted by exposure is to consider that each exposure gives rise to an antibody response. This can be thought of as a set of antibody-producing cells that are triggered by the exposure. Suppose the random variable \( v \) represents the number of such sets of cells, and \( x_i \) is the probability \( P(v = i) \). For every exposure the value of \( v \) has a one-unit increase, and when any of the sets of cells dies the value of \( v \) has a one-unit decrease. If the average duration of a set of cells is \( 1/r \), the rate for the transition from \( v = i \) to \( v = i-1 \) is the product of \( i \) and \( r \) (i.e. \( i x_r \)). Because there are \( i \) sets of cells the value of \( v \) reverts from \( i \) to \( i-1 \) when any of the \( i \) sets of cells dies off (see Figure 1).

The compartmental model (Figure 1) can be represented in differential equation as shown in equation (1) below.

\[
\frac{dx_i}{da} = -\lambda x_i + \lambda x_{i-1} - i r x_i + (i + 1) r x_{i+1}, i = 0, \ldots, \infty; x_i = 0 \text{ for } i < 0
\]

where \( x_i \) is the annual rate of reversion from seropositive to seronegative state per exposure. It follows that

\[
S_{\infty} = \text{Probability (number of exposures, } v = i), \text{ so that } \sum_{i=0}^{\infty} x_i = 1.
\]

These equations can be solved by a standard method, using generating functions [11].
Estimation of parameters:
In the specification of the basic superinfection model there are two main parameters ($\lambda$, $r$). But in the specification of the superinfection model, which allowed an abrupt change in seroconversion rate there are three main parameters ($\lambda_1$, $\lambda_2$, $r$). The model was fitted to age-stratified serological data using the method of maximum likelihood. In this model, the dependent variable $Y$ is an indicator variable, meaning that it takes on only the values 0 (= seronegative state) or 1 (= seropositive state). The probability mass function from the Bernoulli distribution - the distribution for the random indicator variable - is:

$$f(y_j; \pi_j) = \begin{cases} \pi_j & \text{if } y_j = 1 \\ 1-\pi_j & \text{if } y_j = 0 \end{cases}$$

where $0 \leq \pi_j \leq 1$

and $\pi_j$ was identified as the probability for a success (arbitrarily $y_j = 1$ is called a success).

The log-likelihood function for the $j^{th}$ observation is:

$$\ln \ell_j = \begin{cases} \ln(P(a_j)) & \text{if } y_j = 1 \\ \ln(1-P(a_j)) & \text{if } y_j = 0 \end{cases}$$

investigate abrupt changes in malaria transmission in the recent time past.

Figure 1: Compartmental superinfection model.

Figure 2: Comparison of superinfection estimates and catalytic estimates for each village, original data used by Drakeley et al., 2005, Tanzania.
where $y_j$ is the indicator variable $y_j = 1$ if person $j$ is seropositive and $y_j = 0$ if they are seronegative, and $a_j$ is their age and $P(a_j)$ is the proportion seropositive at age $a_j$.

The antibody titre was used to classify individuals as seropositive (or responder) or seronegative (or non-responder) using the mixture model method \[12\]. Briefly, the distribution of normalized optical density (OD) values was fitted as the sum of two Gaussian distributions - a narrow distribution of seronegatives and a broader distribution of seropositives - using maximum likelihood methods. The mean OD of the Gaussian corresponding to the seronegative population plus three standard deviations was used as the cut-off for seropositivity \[6\]. A separate cut off was generated for each antigen, say MSP1 & AMA1.

The seroconversion rate (SCR) was then estimated by fitting the superinfection model to the observed seroprevalence data, stratified into yearly age groups, using the maximum likelihood methods. All members of the population

### Table 1 Comparison of estimates of seroconversion rate (SCR) and duration of seropositivity to MSP1\textsubscript{19} using catalytic and superinfection models fitted to data from 12 villages/sites in Tanzania

| Villages   | Altitude (meters) | $\%$ Catalytic model [5] (1) | $\%$ Superinfection model (2) | $\%$ change in SCR | % change in Std. Err. | Predicted EIR [18] |
|------------|-------------------|------------------------------|------------------------------|---------------------|------------------------|-------------------|
| Mgila      | 375               | 0.1228 (0.0139)              | 0.1451 (0.0116)              | 18                  | -17                    | 39.1002           |
| Kadando    | 528               | 0.0959 (0.0104)              | 0.1097 (0.0098)              | 14                  | -6                     | 16.3467           |
| Kambiri Simba | 746           | 0.0753 (0.0067)              | 0.0857 (0.0073)              | 14                  | 6                      | 4.7182            |
| Ngulu      | 832               | 0.0869 (0.0077)              | 0.0980 (0.0080)              | 13                  | 4                      | 2.8899            |
| Tamota     | 1055              | 0.0574 (0.0061)              | 0.0726 (0.0066)              | 26                  | 8                      | 0.8107            |
| Goha       | 1163              | 0.0239 (0.0028)              | 0.0306 (0.0034)              | 28                  | 24                     | 0.438             |
| Lambo      | 1188              | 0.0059 (0.0017)              | 0.0123 (0.0021)              | 25                  | 24                     | 0.3799            |
| Funta      | 1240              | 0.1033 (0.0108)              | 0.1197 (0.0101)              | 16                  | -7                     | 0.2824            |
| Mpinji     | 1445              | 0.0065 (0.0012)              | 0.0083 (0.0015)              | 27                  | 29                     | 0.0878            |
| Kilomeni   | 1536              | 0.0046 (0.0010)              | 0.0058 (0.0013)              | 27                  | 27                     | 0.0466            |
| Kwadoe     | 1564              | 0.0071 (0.0015)              | 0.0092 (0.0019)              | 29                  | 29                     | 0.0445            |
| Bwambo     | 1598              | 0.0041 (0.0009)              | 0.0054 (0.0012)              | 31                  | 30                     | 0.0367            |
| Average    |                   |                              |                              |                     |                        |                   |
| Reversion rate | 0.0139 (0.0029) | 0.0426 (0.0062)              |                              |                     |                        |                   |
| Duration in years | 72              |                              |                              |                     |                        |                   |

\[ The models were estimated using Stata software's \texttt{ml} command; Standard errors were approximated to 4 decimal places but actual values were used in calculating the % change.\n
![Figure 3](http://www.malariajournal.com/content/13/1/12)

**Figure 3** Association between altitudes and annual rate of seroconversion from MSP1\textsubscript{19} seronegative to seropositive (Superinfection & Catalytic models) or predicted EIR (Bodker et al., 2003) for the 12 sites in Tanzania.
become susceptible at a certain age. This is the age when maternal malaria immunity has sufficiently waned to make the children susceptible to infection. This age is an unknown constant, which can be dealt with using several approaches. Remme et al. [4] estimated this age as part of the model specification. Drakeley et al. [5] excluded children below one year and fitted the model to all villages simultaneously, allowing SCR to vary among villages but with the reversion rate constrained to a single value. Drakeley’s approach has been adopted to fit the superinfection model. The maximum likelihood estimation of the SCR has been executed using the ‘ml’ command in Stata for Windows (College Stations, Texas, USA). Briefly, the log-likelihood function of proportion seropositive at age, \( a \), and its first derivatives was evaluated using the Method-d1 evaluators [13]. Ado-files were then written to maximise the likelihood and predict the SCR using the observed age-stratified seroprevalence data.

**Results**

The superinfection model was compared with the original data from Tanzania. The results showed that estimates of seroconversion rate were higher using the superinfection model than catalytic model (Figure 2). This difference was

![Graphs showing seropositivity in North Pare villages](image1)

**Figure 4** Superinfection model of seropositivity to MSP119 for villages in North Pare, Tanzania.

![Graphs showing seropositivity in South Pare villages](image2)

**Figure 5** Superinfection model of sero-positivity to MSP119 for villages in South Pare, Tanzania.
milder when the level of transmission was lower. This sug-
gests that the catalytic model might be underestimating
the transmission intensity by up to 31\% (Table 1). The
duration is shorter with superinfection model, but still
seems too long (Table 1). This may be because the model
assumes a constant reversion rate. The estimates of
seroconversion rate were lower than the predicted ento-
mological inoculation rate (Figure 3). It may be that the
predictive model is no longer valid. The seroprevalence
profiles appeared to suggest that the model fits the sero-
conversion rate quite well but it may be difficult to discriminate
the model on the basis of fit alone (Figures 4, 5 and 6).
The superinfection model was also fitted to data from
Bioko (Table 2) where a change in transmission has been
established using the catalytic model. The superinfection
model predicts a change in transmission, which is consist-
ent with the catalytic model (Figure 7 and Table 3). The
model, which allowed an abrupt change in seroconversion
rate were better fit than the model, which did not assume
that seroconversion rate has changed (Likelihood ratio test
for AMA1: LR chi2 (1) = 60.11; P-value < 0.0001; and for
MSP1\textsubscript{19}: LR chi2 (1) = 3.09; P-value = 0.08).

Discussion
In this paper, a mathematical model of seropositivity to
malaria antigens that allows seropositivity to be prolonged
by exposure was developed. This model is important
because it can produce more realistic estimates of the
duration of seropositivity. This is analogous to Dietz model,
which allowed for superinfection and produced more realis-
tic estimates of the duration of infection as compared to
the original Ross-MacDonald malaria model [14-17], which
also ignores superinfection. However, some discrepancies
may remain, considering the simplifying assumptions in-
cluded in the model. In particular, the reversion rate may
depend on age and the seroconversion rate may also
depend on age.

The estimates of seroconversion rates at the 12 villages in
Tanzania were lower than Bodker and colleague's predicted
EIR [18]. It may be that the predictive model developed
by Bodker \textit{et al.} [18] is no longer valid: the dynamics of
malaria transmission might have changed and the data
used for the predictive model was collected a decade
(1995/96) before the current study by Drakeley \textit{et al.},
to which the model is being applied. The estimate of the
mean duration of seropositivity was lower in the super-
infection model than the catalytic model but it still seems
too long. One possible reason is that the model assumed a
constant reversion rate but it may depend on age. For ex-
ample, in a cohort study of young children, aged six years

| Age group (years) | Seropositive to AMA1 | Seropositive to MSP1\textsubscript{19} |
|------------------|----------------------|--------------------------|
| Number of        | Number of            |                          |
| participants     | participants         |                          |
| %                | %                    |                          |
| <1               | 611 13.9             | 615 6.5                  |
| 1 <2             | 292 13.4             | 300 8                    |
| 2 <5             | 720 22.6             | 728 13.3                 |
| 5 <10            | 938 45               | 934 13.9                 |
| 10 <15           | 617 61.8             | 621 21.4                 |
| 15 <20           | 275 80.7             | 269 33.5                 |
| 20 <25           | 349 78.8             | 358 39.1                 |
| 25 <35           | 568 76.9             | 566 43.6                 |
| 35 <50           | 650 69.1             | 635 41.9                 |
| 50+              | 533 65.3             | 528 44.9                 |
| Total            | 5553 50.8            | 5554 25.3                |

Figure 6 Superinfection model of sero-positivity to MSP1\textsubscript{19} for villages in West Usambara, Tanzania.
and below, conducted in The Gambia to investigate the determinants of antibody response longevity, Akpogheneta and colleagues have demonstrated that antibodies decayed more slowly among children in the oldest age group and more rapidly among children in the youngest age group [19]. The superinfection model also assumes a constant seroconversion rate but it may also depend on age. For example, Nasell has shown that young children may experience superinfection up to age 5 but it is not common in adults [10]. In a multistrain model, Milligan and Downham found similar conclusions that the fraction of infection increased more rapidly in younger children than in adults as they acquire immunity to different strains of pathogens [9]. In a recent study, Portugal and colleagues held an opposite view that superinfections are uncommon in younger children [20]. A model that allowed the mean duration of seropositivity to depend on age may give a more realistic estimate of the seroconversion rate.

A better fit could probably be obtained only at the cost of complex model specification. Future work should consider extending the model to allow the reversion rate to depend on age according to an exponential function. The function can be chosen such that the duration of seropositivity is three years for adults, which is consistent with results from longitudinal studies. A model that allowed the seroconversion rate to drop gradually according to a logistic function is also an important consideration. This may be a more realistic model for how seroconversion rate changes over time than the model that assumes an abrupt change. It may not be possible to fit complex models with age-dependent duration, the parameters may not be identifiable. It may be necessary to consider the mean duration (as a function of age) to be fixed, determined from field data, and then use the superinfection model with the duration parameters fixed, to estimate the seroconversion rate, perhaps with the more flexible logistic function used to estimate trends over time. This then is a topic for further research, to find out if the mean duration can be considered constant for a particular type of antigen and it would be necessary to find reliable estimates.

**Table 3 Parameter estimates for the superinfection model, Bioko, 2008**

| Parameters | AMA1 | MSP1 |
|-----------|------|------|
| λ₁        | 0.611| 0.071|
| λ₂        | 0.04 | 0.022|
| r         | 0.178| 0.072|

λ₁ = Previous seroconversion rate; λ₂ = Current seroconversion rate; r = reversion rate.

**Figure 7** Log-likelihood profiles and superinfection models of probability seropositive to AMA1 & MSP1, allowing abrupt change in incidence rate (Rabola, Bakake, Baney), Bioko, 2008.
Conclusion
The superinfection model is important because it can produce more realistic estimates of the duration of seropositivity. This is analogous to Dietz model, which allowed for superinfection and produced more realistic estimates of the duration of infection as compared to the original Ross-MacDonald malaria model, which also ignores superinfection.

Competing interests
The author declares that he has no competing interests.

Acknowledgements
I am very thankful to Paul Milligan, PhD, Reader in Epidemiology and Medical Statistics, London School of Hygiene and Tropical Medicine (LSHTM), for his guidance and mentorship. I would also like to thank Immo Kleinschmidt, PhD, Reader in Medical Statistics and Epidemiology, Chris Drakeley, PhD, Senior Lecturer and Director of LSHTM Malaria Centre, Lucy Okel, PhD, Research Fellow and Jackie Cook of LSHTM for generously providing field data for the model validation.

Received: 2 September 2013 Accepted: 3 January 2014
Published: 8 January 2014

References
1. Muench H. Catalytic models in epidemiology. Cambridge, Mass: Harvard University Press; 1959.
2. Anderson RM, May RM: Directly transmitted infections diseases: control by vaccination. Science 1982, 218:1053–1060.
3. Griffiths DA: A catalytic model of infection for measles. Appl Stat 1974, 23:330–339.
4. Remme J, Mandara MP, Leeuwenberg J: The force of measles infection in East Africa. Int J Epidemiol 1984, 13:332–339.
5. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SL, Carneiro I, Malima R, Lusingu J, Manjurano A, Nyika WM, Lemnge MM, Cox J, Reyburn H, Riley EM: Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci U S A 2005, 102:5108–5113.
6. Cook J, Kleinschmidt I, Schwabe C, Nseng G, Bousema T, Corran PH, Riley EM, Drakeley CJ: Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea. PLoS One 2011, 6:e25137.
7. Bailey NTJ: The mathematical theory of epidemics. London: Griffin; 1957:157–159.
8. Bailey NTJ: The biomathematics of malaria. London: Charles Griffin, 1982.
9. Milligan P, Downham DY: Models of superinfection and acquired immunity to multiple parasite strains. J Appl Probab 1996, 33:915–932.
10. Nasell I: On superinfection in malaria. Math Med Biol 1986, 3:211–227.
11. Bailey NTJ: The elements of stochastic processes with applications to the natural sciences. John Wiley & Sons; 1990. ISBN 13: 978-0-471-25368-0.
12. Stewart L, Gosling R, Griffin J, Gerase S, Campo J, Hashim R, Masika P, Mosha J, Bousema T, Shekalaghe S, Cook J, Corran P, Ghani A, Riley EM, Drakeley C: Rapid assessment of malaria transmission using age-specific sero-conversion rates. PLoS One 2009, 4:e6683.
13. Gould W, Piblado J, Poi B: Maximum likelihood estimation with stata. 4th edition. 4005 Lualenyi Drive, College Station, Texas 77485. Stata Press; 2010.
14. Macdonald G: The analysis of the sporozoite rate. Trop Dis Bull 1952, 49:569–586.
15. Macdonald G: Epidemiological basis of malaria control. Bull World Health Organ 1956, 18:513–626.
16. Macdonald G: The epidemiology and control of malaria. London: Oxford University Press; 1957.
17. Ross R: The prevention of malaria. 2nd edition. London: Murray; 1911.
18. Badker R, Akida J, Shayo D, Kisirna W, Msangi HA, Pedersen EM, Lindsay SW: Relationship between altitude and intensity of malaria transmission in the Usambara Mountains, Tanzania. J Med Entomol 2003, 40:706–717.
19. Akpogheneta OJ, Duah NO, Tetteh KKA, Dunyo S, Lanar DE, Pinder M, Conway DJ: Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. Infect Immun 2008, 76:1748–1755.
20. Portugal S, Carret C, Recker M, Armitage AE, Gonçalves LA, Epiphanio S, Sullivan D, Roy C, Newbold CI, Drakesmith H, Mota MM: Host-mediated regulation of superinfection in malaria. Nat Med 2011, 17:732–737.