Notes on reference datasets and simulation assumptions used in validation

Updated May 12, 2022

### Chonyi, Kenya, 1999-2001

[Paper: Mwangi et al. 2005](https://academic.oup.com/jid/article/191/11/1932/2192008?login=false).

#### *Case management*

In the study, each household was visited once a week, with temperature taken. Participants with fever were given bus fare to travel to study clinic to get a blood smear and receive treatment. For participants who could not go to the study clinic, blood smear was performed in the field and if participant was positive, the worker provided malaria treatment on the following day. First line treatment was sulfadoxine-pyrimethamine throughout the study. Estimates of treatment rates were not reported, so **we assume a coverage of 70% for participants with symptomatic malaria but note that this estimate is very uncertain**.

The case management before the beginning of the study is not specified but could be important in influencing the amount of immunity in the population. We assume a coverage of 30% before the beginning of the study for lack of a more informed estimate.

#### Case detection rate

Cases were detected through weekly household visits from active surveillance as well as from passive surveillance (passive surveillance not described, but presumably from local study clinic or health facilities). The paper does not report how often participants were absent when the household visit occurred, making it difficult to estimate observation rate. **We assume that 90% of febrile cases would have been detected, though note that this is quite uncertain**.

#### *EIR*

[Mbogo et al. 2003](https://www.ajtmh.org/view/journals/tpmd/68/6/article-p734.xml) looked at vector counts and sporozoite rates in numerous sites to calculate EIR and show vector densities over a year. In Kilifi district, overall mean daily human-biting-rate for *An. gambaie* was 0.43 and sporozoite rate was 2.88, with a mean daily entomologic inoculation rate of 0.043 (annual EIR of 16). Across sites with more than 100 mosquitoes used to calculate sporozoite rate, the daily EIR ranged from 0.019 to 0.116, giving an annual EIR between 7 and 42. For *An. funestus* in Kilifi district, overall mean daily human-biting-rate was 0.23 and sporozoite rate was 4.89, with a mean daily entomologic inoculation rate of 0.010 (annual EIR of 3.7). Across sites with more than 100 mosquitoes used to calculate sporozoite rate, the daily EIR ranged from 0.018 to 0.023, giving an annual EIR around 7. Combining *An. gambaie* and *An. funestus*, in sites with more than 100 mosquitoes sampled, the annual EIR ranged from 13.5 to 52.9, with an aggregate across sites of 19.3. Some of these sites may be further from Chonyi than others. For example, from the maps shown in the publications, it appears that Barani, which had the lower bounds, is further from Chonyi than the other sites, so the lower range among the closest sites may be closer to 33. Mwangi et al. 2005 cites Mbogo et al. 2003 as giving the Chonyi EIR as 22-53 bites/person/year. **We assume that annual EIR is 35 but note considerable uncertainty in this assumption.**

~~To get monthly EIR, we used mean monthly mosquito density in Kilifi to rescale annual EIR for each month. We do not expect that this will be a perfect match to the true monthly EIRs – for instance, there is likely a delay between when mosquito populations increase and when sporozoite rates increase. However, lacking better data, hopefully it will be a close enough approximation to capture the impact of seasonality in the site. The monthly densities were extracted from Figure 3 in Mogo et al. 2003 (row for Kilifi), adding the values for~~ *~~gambiae~~* ~~and~~ *~~funestus~~*~~.~~

Instead of using monthly mosquito density in Kilifi from Mgobo et al. 2003 to rescale annual EIR for monthly EIR, we rescale annual EIR with the relative EIRs in each month reported in Mbogo et al. 1995 Table 3, taking the average across nine sites a bit north of Chonyi. See Mbogo\_1995\_Kenya\_EIR\_seasonality\_Table3.xlsx.

#### Reference datasets

Age-incidence and age-prevalence reference datasets come from the csv associated with [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21).

For the prevalence surveys, there were three conducted during the wet season (July 1999, 2000, 2021) and three during the dry season (March 2000, October 2000, and March 2001). The paper is not completely explicit in specifying that the participants in the cross-sectional surveys were the same as followed longitudinally for incidence, but it appears to be the case. The recorded prevalence by age group in the Cameron dataset has the month ‘July,’ but I am uncertain whether it includes only the July surveys or might be the average across all six surveys. If the latter, to best match the sampling for prevalence in the simulations, the same months should be averaged for each age group rather than using the annual prevalence. But for now we assume it is the July-survey only. In calculating likelihoods for recalibration, might be reasonable to approximate the population sizes with the population sizes in each age group from the longitudinal incidence records.

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 27 years are used as a burnin (to allow for acquisition of immunity), with low case management and the final three years are meant to mimic the study with higher treatment. Reports from the final years are generated and used to compare with reference datasets.

### Ngerenya, Kenya, 1999-2001

[Paper: Mwangi et al. 2005](https://academic.oup.com/jid/article/191/11/1932/2192008?login=false).

Case management

(Same as for Chonyi site.) Each household was visited once a week, with temperature taken. Participants with fever were given bus fare to travel to study clinic to get a blood smear and receive treatment. For participants who could not go to the study clinic, blood smear was performed in the field and if participant was positive, the worker provided malaria treatment on the following day. First line treatment was sulfadoxine-pyrimethamine throughout the study. Estimates of treatment rates were not reported, so **we assume a coverage of 70% for participants with symptomatic malaria but note that this estimate is very uncertain**.

The case management before the beginning of the study is not specified but could be important in influencing the amount of immunity in the population. We assume a coverage of 30% before the beginning of the study for lack of a more informed estimate.

#### Case detection rate

(Same as for Chonyi site.) Cases were detected through weekly household visits from active surveillance as well as from passive surveillance (passive surveillance not described, but presumably from local study clinic or health facilities). The paper does not report how often participants were absent when the household visit occurred, making it difficult to estimate observation rate. **We assume that 90% of febrile cases would have been detected, though note that this is quite uncertain.**

#### EIR

[Mbogo et al. 1995](https://pubmed.ncbi.nlm.nih.gov/7694959/) sampled mosquitoes over one year in areas near the Ngerenya site (from the maps, it appears that the Mukombe (MU), Kambi ya Wari (KW), and Kaoyeni (KY) sites were closest). They estimated annual bites per person and fraction of mosquitos positive for *P. falciparum*. The annual EIRs estimated for the three closest sites ranged from 2.5 to 3.8 and for all study sites ranged from 0 to 59.6. Mbogo et al. 1995 as giving the Ngerenya EIR as 10 bites/person/year.

Monthly EIRs were also calculated for each site and reported in Table 3. Since the numbers were small and sparse for the three closest sites, we use the average fraction of EIR in each month across all nine sites (such that each site is weighted equally, regardless of annual EIR). We then multiply these monthly fractions by the average annual EIR of the three closest sites (3.3) to get the monthly EIRs. See Mbogo\_1995\_Kenya\_EIR\_seasonality\_Table3.xlsx.

#### Reference datasets

Age-incidence and age-prevalence reference datasets come from the csv associated with [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21).

For the prevalence surveys, there were three conducted during the wet season (July 1999, 2000, 2021) and three during the dry season (March 2000, October 2000, and March 2001). The paper is not completely explicit in specifying that the participants in the cross-sectional surveys were the same as followed longitudinally for incidence, but it appears to be the case. The recorded prevalence by age group in the Cameron dataset has the month ‘July,’ but I am uncertain whether it includes only the July surveys or might be the average across all six surveys. If the latter, to best match the sampling for prevalence in the simulations, the same months should be averaged for each age group rather than using the annual prevalence. But for now we assume it is the July-survey only. In calculating likelihoods for recalibration, might be reasonable to approximate the population sizes with the population sizes in each age group from the longitudinal incidence records.

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 27 years are used as a burnin (to allow for acquisition of immunity), with low case management and the final three years are meant to mimic the study with higher treatment. Reports from the final years are generated and used to compare with reference datasets.

### Dielmo, Senegal, 1990-1993

References: [Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/), [Trape et al. 1994](https://pubmed.ncbi.nlm.nih.gov/8074247/), [Rogier and Trape 1995](https://www.documentation.ird.fr/hor/fdi:010005112)

#### Case management

Each villager was visited at home three times a week to have their temperature taken and report any symptoms within the prior 48 hours according to [Trape et al. 1994](https://pubmed.ncbi.nlm.nih.gov/8074247/). (Note that in [Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/), the frequency of home visits is described as daily, so it is possible it changed over time). Twice a week, a thick blood film was made. In addition, each compound was visited daily to detect new cases of fever and the dispensary was open at all times. All individuals reporting fever had a thick blood film examined and was treated with quinine if determined positive. The positive determination varied by age and situation: febrile individuals with a parasite:leukocite ratio of 2 or greater in children or 0.5 or greater in pregnant women, or any detected parasites in individuals who had recently returned from an area with low malaria were assumed positive and treated. For remaining patients with symptoms, condition would be monitored and under several conditions treatment would be given for malaria in subsequent days (described in [Trape et al. 1994](https://pubmed.ncbi.nlm.nih.gov/8074247/)).

This high rate of treatment was not present before the study, it may be preferable to simulate burnins without treatment and only include the high rate of treatment during the study year rather than simulating a single cohort with consistent high treatment. In [McCarthy et al. 2015](https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-14-6), there was a 30% probability per day of treatment after fever onset, which would mean that nearly everyone was treated within a few days. In [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y), it was assumed that 30% of symptomatic cases received curative treatment in total (with delays in treatment not specified). We assume that 90% of individuals with clinical malaria symptoms received treatment.

#### Case detection rate

Given the frequency of monitoring, we assume that 95% of cases were recorded.

#### EIR

Taken from prior recalibration setup (I assume – values were in [input\_eir\_by\_site.py](https://github.com/InstituteforDiseaseModeling/idmtools-calibra/blob/0d8aa8f6ef867fe9d2823dc9536eceec3759bdd7/malaria/site/input_eir_by_site.py#L37)). I assume these values are the same ones used in [McCarthy et al. 2015](https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-14-6) and [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y).

For future reference, the estimated monthly EIRs from 1990-1992 are reported in Figure 4 of [Trape et al. 1994](https://pubmed.ncbi.nlm.nih.gov/8074247/) and could be extracted. [Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/) estimates the Dielmo EIR as around 200 infective bites/person/year.

#### Reference datasets

Age-incidence and age-prevalence reference datasets come from the csv associated with [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21).

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 27 years are used as a burnin (to allow for acquisition of immunity), with only severe case management and the final three years are meant to mimic the study with higher treatment. Reports from the final years are generated and used to compare with reference datasets.

~~Note that we simulate treatment with chloroquine instead of quinine, since quinine is not currently a default drug (~~[~~Watt et al 1988~~](https://pubmed.ncbi.nlm.nih.gov/3055454/) ~~suggest that the efficacies of the two drugs are similar).~~ Originally, we simulated chloroquine instead of quinine, but this yielded very long protection against reinfection. We now assume the same parameters as ACT for quinine. See [White 2021](https://malariajournal.biomedcentral.com/articles/10.1186/s12936-021-03700-7) Figure 1 for temporal distribution of *P. vivax* relapse times with different drugs. The clinical incidence results are hugely sensitive to the choice of drug parameters.

Note: the previous calibration approach was to assume a constant rate of case management throughout an individual’s life, we now run simulations with vital dynamics and a burnin, followed by the study years of treatment at 90%.

### Ndiop, Senegal, 1993-1994

References: [Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/), [Trape et al. 1994](https://pubmed.ncbi.nlm.nih.gov/8074247/), [Rogier and Trape 1995](https://www.documentation.ird.fr/hor/fdi:010005112)

#### Case management

(Similar to Dielmo 1990-1994.) Each villager was visited at home daily to have their temperature taken and report any symptoms, with any “pathological” episodes diagnosed and treated ([Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/)), Twice a week, a thick blood film was made. All individuals reporting fever had a thick blood film examined and was treated with quinine if determined positive. The positive determination varied by age and situation, with threshold parasite:leukocite ratio of 2.45 for children of age one year to 0.5 for individuals over 60 years of age.

(Same as Dielmo 1990-1994.) This high rate of treatment was not present before the study, it may be preferable to simulate burnins without treatment and only include the high rate of treatment during the study year rather than simulating a single cohort with consistent high treatment. In [McCarthy et al. 2015](https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-14-6), there was a 30% probability per day of treatment after fever onset, which would mean that nearly everyone was treated within a few days. In [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y), it was assumed that 30% of symptomatic cases received curative treatment in total (with delays in treatment not specified).

(Same as Dielmo 1990-1994.) The currently coded approach is to assume a constant rate of case management throughout an individual’s life, but I think this should be adjusted to run simulations with vital dynamics and a burnin, followed by the study years of treatment at 90%. Also should add the summary report start day as when the ‘trial’ begins as the simulation instead of from the beginning. And increase population size or have more seeds.

#### Case detection rate

Given the frequency of monitoring, we assume that 95% of cases were recorded.

#### EIR

Taken from prior recalibration setup (I assume – values were in [input\_eir\_by\_site.py](https://github.com/InstituteforDiseaseModeling/idmtools-calibra/blob/0d8aa8f6ef867fe9d2823dc9536eceec3759bdd7/malaria/site/input_eir_by_site.py#L37)). I assume these values are the same ones used in [McCarthy et al. 2015](https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-14-6) and [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y).

[Rogier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10697865/) estimates the Ndiop EIR as around 20 infective bites/person/year and describes that transmission only occurs during the four months of the rainy season.

#### Reference datasets

Age-incidence and age-prevalence reference datasets come from the csv associated with [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21).

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 27 years are used as a burnin (to allow for acquisition of immunity), with only severe case management and the final three years are meant to mimic the study with higher treatment. Reports from the final years are generated and used to compare with reference datasets.

~~Note that we simulate treatment with chloroquine instead of quinine, since quinine is not currently a default drug (~~[~~Watt et al 1988~~](https://pubmed.ncbi.nlm.nih.gov/3055454/) ~~suggest that the efficacies of the two drugs are similar).~~ Originally, we simulated chloroquine instead of quinine, but this yielded very long protection against reinfection. We now assume the same parameters as ACT for quinine. See [White 2021](https://malariajournal.biomedcentral.com/articles/10.1186/s12936-021-03700-7) Figure 1 for temporal distribution of *P. vivax* relapse times with different drugs. The clinical incidence results are hugely sensitive to the choice of drug parameters.

Note: the previous calibration approach was to assume a constant rate of case management throughout an individual’s life, we now run simulations with vital dynamics and a burnin, followed by the study years of treatment at 90%.

## Cameroon 1997-1998 sites

### Ebolakounou, Cameroon, 1997-1998

### Koundou, Cameroon, 1997-1998

Paper: [Bonnet et al. 2002](https://pubmed.ncbi.nlm.nih.gov/11903987/)

#### Case management

Trained field workers visited participants’ homes daily and took blood films from all individuals who were suspected of having malaria. Confirmed malaria cases were treated with amodiaquine (30mg/lg for 3 days). The health workers took new blood slides 7 and 14 days after treatment and found 23% were positive (which could be from recrudescence or from new infections). We assume that 90% of individuals with clinical malaria symptoms received treatment during the study and that only severe cases were treated before the study onset.

#### Case detection rate

Given the daily sampling, we assume that 95% of clinical cases of malaria were detected.

#### EIR

The year-long entomological study reported in Bonnet et al. (originally published in [Meunier et al. 1999](https://pubmed.ncbi.nlm.nih.gov/10690465/)), gives 12.7 bites/person/night in Koundou and 1.1 bites/person/night in Ebolakounou. The calculate the EIR as 176.1 in Koundou and 17.7 in Ebolakounou. However, the study only looked at ten houses twice a month, so estimated EIR is zero in a number of months in both sites (especially in Ebolakounou). Figures 2-3 of Meunier et al. 1999 show EIRs for certain days when sampling occurred. We started by calculating the mean of the two sample days in each month for each site, though for Ebolakounou especially, the numbers are small enough that we decided to assume that Ebolakounou has the same seasonality as seen in Koundou. Ultimately, to calculate the monthly EIR, we 1) calculate the mean of the two sample dates in each month for Koundou, 2) set a monthly minimum EIR of 0.1, 3) rescale the monthly values so that the total equals the reported annual EIR for the site.

#### Reference datasets

Age-incidence and age-prevalence reference datasets come from the csv associated with [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21).

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 27 years are used as a burnin (to allow for acquisition of immunity), with only severe case management and the final three years are meant to mimic the study with higher treatment. Reports from the final years are generated and used to compare with reference datasets.

## Mali 1999-2000 sites

### Dongubougou, Mali, 1999-2000

### Sotuba, Mali, 1999-2000

Paper: [Dicko et al. 2007](https://www.ajtmh.org/view/journals/tpmd/77/6/article-p1028.xml)

#### Case management

There are medical clinics in Doneguebougou and Sotuba maintained by the Malaria Research and Training Center of the University of Bamako.

The study included weekly visits during the malaria transmission season, with fingerprick blood samples taken every four weeks. If individuals had malaria symptoms (at the weekly or every-four-month visits), blood was taken and examined immediately. People who had malaria symptoms and any asexual parasite density were treated with a single weight-adjusted dose of sulfadoxine-pyrimethamine and severe cases were treated with quinine.

The Malaria Research and Training Center was created in 1992 (according to a brief internet search), so we assume that case management rates were near zero prior to 1992, 50% following 1992, and 80% during the study.

#### Case detection rate

Cases were detected during weekly (active surveillance) visits or through passive surveillance when individuals visited the health clinics. Around 52% of cases were detected in the health clinics and the remainder of cases were detected during the scheduled visits, with similar ratios across ages.

We assume that 90% of cases were detected.

#### EIR

Dicko et al. reports that “There is essentially no malaria transmission at either site during the dry season (January to June) (Y. T. Toure, unpublished results).”

Between June to December, monthly human landing catches (HLC) were done in two locations in each village, both inside and outdoors, and circumsporozoite antigen detection rates were calculated, which combined to give monthly EIR estimates. Both PSC and HLC were used to calculate EIR, giving wildly different estimates (Table 2 of Dicko et al. 2007). We use the monthly estimates generated with HLCs, taking the average between the two surveyed years. For non-surveyed months or months with a reported EIR of 0, we assume the EIR is 0.01.

#### Reference datasets

The reference data for age-incidence and age-prevalence for these sites come from the [Cameron et al. 2015](https://www.nature.com/articles/ncomms9170#Sec21) dataset.

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years: the first 20 years are used as a burnin (to allow for acquisition of immunity), with only severe case management, the next 7 years have moderate levels of case management, and the final three years are meant to mimic the study with high treatment rates. Reports from the final years are generated and used to compare with reference datasets.

### Nanoro, Burkina Faso

*Coming soon…*

### Navrongo, Ghana, 2000

Navrongo, Ghana site (match study ages) – relevant papers include: [Bretscher et al. 2011](https://www.sciencedirect.com/science/article/pii/S175543651100020X?via%3Dihub), [Bretscher et al. 2015](https://pubmed.ncbi.nlm.nih.gov/26238109/), [Felger et al. 2012](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045542), [Owusu-Aguei et al. 2002](https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-3156.2002.00881.x?sid=nlm%3Apubmed)

[Data on Malaria Team Dropbox](https://www.dropbox.com/home/Malaria%20Team%20Folder/data/Ghana/Navrongo)

#### Case management

Treatment was not administered as part of the study, though individuals who were sick at a survey date “were referred to routine health services” ([Felger et al. 2012](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045542)). It does not sound like antimalarial treatments were common, and most were likely given to younger children: “For the youngest age groups, short durations may partly reflect effects of antimalaria treatment, mainly affecting <2 years old children. Treatment was infrequent but could not be formally incorporated into the models because of absence of comprehensive individual-level data” ([Felger et al. 2012](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0045542)).

Lacking other information, we assume a treatment rate for clinical malaria of 20% in children U5, 10% in adults, and 50% in severe cases, both before and during the study period.

#### EIR

The average of three years (2001-2003) of monthly EIRs reported in Kasasa et al. 2013 were used. The resulting seasonality and magnitude of EIRs are fairly similar to the EIRs used for Dapelogo, which is somewhat nearby. However, there was a substantial amount of variation between the years in Kasasa et al, so adjustments to the magnitude of EIR may be warranted.

#### Reference datasets

The parasite density and duration-of-infection datasets come from an age-stratified cohort of 347 individuals who were sampled every two months for a year (for a total of six samples each).

#### Simulations

Simulations with demographic vital dynamics are run for a total of 30 years with constant, relatively low case management rates. The first 27 years are used as a burnin (to allow for acquisition of immunity), and simulation output from the final three years are saved and compared against reference datasets.

## Ouedraogo’s 2007 Burkina Faso sites

Papers: [Ouedraogo et al. 2016](https://academic.oup.com/jid/article/213/1/90/2459230?login=true), also information in EMOD calibration paper [Gerardin et al. 2015](https://malariajournal.biomedcentral.com/articles/10.1186/s12936-015-0751-y)

### Laye, Burkina Faso, 2007-2008

### Dapelogo, Burkina Faso, 2007-2008

#### Case management

From the Burkina Faso HBHI work, the effective treatment rates for clinical malaria appeared to be quite low prior to 2010 (as seen in the DHS results). We assumed that 10% of children and adults with clinical symptoms received effective treatment and 60% of individuals with severe malaria received treatment.

#### EIR

Taken from prior recalibration setup (I assume – values were in [input\_eir\_by\_site.py](https://github.com/InstituteforDiseaseModeling/idmtools-calibra/blob/0d8aa8f6ef867fe9d2823dc9536eceec3759bdd7/malaria/site/input_eir_by_site.py#L37)). As reported in [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y): “In the Burkina Faso sites, modeled seasonality followed entomological observations, with an annual EIR of 300 in Dapelogo and 30 in Laye [[33](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y#ref-CR33)]. ”

#### Reference data

The longitudinal study [Ouedraogo et al. 2016](https://academic.oup.com/jid/article/213/1/90/2459230?login=true) measured asexual and gametocyte densities and used membrane feeding tests to quantify infectiousness to mosquitoes. The data values were taken from the hardcoded values in [prior recalibration scripts](https://github.com/InstituteforDiseaseModeling/idmtools-calibra/tree/0d8aa8f6ef867fe9d2823dc9536eceec3759bdd7/malaria/study_sites).

#### Simulations

A cohort of individuals is simulated for 65 years, with simulation results saved throughout the simulation. There are no births or deaths. The case management rates are assumed to be constant and low throughout the simulation.

## The Garki Project

Paper: Molineaux et al. 1980. The Garki project.

### Matsari, Nigeria, 1970-1972

### Rafin Marke, Nigeria, 1970-1972

### Sugungum, Nigeria, 1970-1972

#### EIR

Taken from prior recalibration setup (I assume – values were in [input\_eir\_by\_site.py](https://github.com/InstituteforDiseaseModeling/idmtools-calibra/blob/0d8aa8f6ef867fe9d2823dc9536eceec3759bdd7/malaria/site/input_eir_by_site.py#L37)). As reported in [Selvaraj et al. 2018](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y): “In the Garki sites, monthly EIR was calculated by averaging the product of human biting rate and sporozoite rate over the days of the month when data was collected [[2](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y#ref-CR2)], then multiplying this value by the number of days in the month. Parasite prevalence data from infants, stratified by density and month of observation, were used to infer monthly EIR when entomological data was not available. Monthly EIR values were adaptively tuned for a given parameter set until there was a good fit of simulation data to the reference dataset (Fig. [1](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-018-3319-y#Fig1)). To address dry-season biting, which appeared to be minimal in infants, the population was divided into two groups: one experiencing the infant-based EIR profile, and another experiencing higher EIR during the dry-season months of March, April, and May.” We do not include the two groups that experience different levels of dry-season biting; they were not specified in the python script with the input EIRs.

#### Case management

In control areas and before the study commenced (in the ‘baseline’ phase between October 1970-March 1972), it sounds as though pharmaceutical treatments for malaria were not common: “Knowledge and utilization of antimalarials in the study area were practically negligible before the research project commenced” (page 50 of Molineaux et al. 1980). These will be the periods included in simulations, so reference data will need to be subset accordingly or control regions will need to be selected to avoid confounding with treatments of MDA and net distribution. In certain groups, SP or chloroquine was given in 1972 and 1973 (I believe these were the MDA groups). In 1974 and 1975, chloroquine was given to clinical cases.

#### Reference data

Comes from [Garki\_df.csv](https://github.com/pselvaraj87/Malaria-GarkiDB/blob/master/Data/Garki_df.csv) in the [Malaria-GarkiDB repo](https://github.com/pselvaraj87/Malaria-GarkiDB/blob/master/Data/Garki_df.csv).

The Garki data gives the fraction of microscopy fields with at least one parasite detected rather than a count of parasites per uL, which is the metric estimated in simulations. To compare simulation output with reference values, we need to translate between the two.

Following comment in [Garki\_population\_summary.py](https://github.com/pselvaraj87/Malaria-GarkiDB/blob/58463eba08cfb5bd95bda783f67dfe3678cd571f/Garki_population_summary.py#L371), to convert from the values given in the Garki dataset to parasites density (parasites per uL), we assume that there were 0.5 uL per 200 views (page 111 in pdf), meaning that each field has 0.5/200 uL. If the parasite density were X parasites per uL, the probability of seeing parasites in N fields out of 200 could be approximated as follows:

* If the total volume of blood is V and the true density is X, then in a sample of blood of volume S, we’d expect to see the number of parasites following a binomial distribution with n=X\*V and p=S/V.
* Since the total volume of blood is substantially greater than the sample size, we can approximate this with a Poisson distribution. Note that this approximation will be pretty good when the density is not tiny, but when the total number of parasites (X\*V) is very small (say…<100), we don’t expect this approximation to do well. Similarly, if X\*S is small, say, <20, we do not expect a good approximation. The new approximate distribution is Poisson with mean S\*X.
* With this approximation, the probability of seeing zero parasites in a field would be exp(-S\*X).
* Since fields are assumed to be independent conditional on the total parasite density, the probability of observing at least one parasite in N out of 200 fields would be (200 choose N) \* (1-exp(-S\*X))^(N) \* (exp(-S\*X))^(200-N).

We can convert the simulation densities into expected fraction of positive fields, with E(N/200|X) = 200\*(1-exp(-S\*X)) / 200 = (1-exp(-0.5/200\*X)). The current parasite density bins included in simulations are {0, 50, 500, 5000, 5000000}, corresponding to fraction of positive fields: {0, 0.1175, 0.7135, 1, 1}.

#### Simulations

A cohort of individuals is simulated for 65 years, with simulation results saved throughout the simulation. There are no births or deaths. The case management rates are assumed to be constant and low throughout the simulation.

## RTS,S Phase III sites

### Kilifi, Kenya, 2009-2014

### Korogwe, Tanzania, 2009-2014

### Manhica, Mozambique, 2009-2014

### Lambarene, Gabon, 2009-2014

### Bagamoyo, Tanzania, 2009-2014

### Lilongwe, Malawi, 2009-2014

### Agogo, Ghana, 2009-2014

### Kombewa, Kenya, 2009-2014

### Kintampo, Ghana, 2009-2014

### Nanoro, Burkina Faso, 2009-2014

### Siaya, Kenya, 2009-2014

## Potential future sites and references

* Manhica, Mozambique, 1996-1999 – incidence/age, prevalence/age
  + Paper: [Saute et al 2003](https://pubmed.ncbi.nlm.nih.gov/16117957/) (paywalled)
* Matola, Mozambique, 1994-1995 – incidence/age, prevalence/age
  + Paper: [Thompson et al. 1997](https://pubmed.ncbi.nlm.nih.gov/9392594/) (paywalled)
* Idete village, Tanzania, 1992-1993 – parasite-density/age, prevalence/age at each month for first year of life
  + Paper: [Kitua et al. 1996](https://pubmed.ncbi.nlm.nih.gov/8765455/)
* High, moderate, low transmission sites in Uganda, 2011-2014 – incidence/age, parasite density/age, parasite density and fever
  + Paper: [Rodriguez-Barraquer et al. 2018](https://elifesciences.org/articles/35832)
* Across many sites, relation between EIR and prevalence
  + Paper: [Smith et al. 2005](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3128496/)
* Ndiop, Senegal, 1993-2013 – parasitemia thresholds for fever, parasite-density/age
  + Paper: [Dollat et al. 2019](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0217903#sec006)