Notes on Open Malaria

Bradley Gram-Hansen

November 29, 2018

1 What is Open Malria?

Open malaria is for simulating malaria epidemiology and the impacts on that epidemiology of interventions against malaria. It is based on microsimulations of Plasmodium falciparum malaria in humans, originally developed for simulating malaria vaccines. These models simulate the dynamics of malaria parasitaemia in the course of an infection, of transmission, of immunity, and of the processes leading to illness and death. The system is set up to simulate malaria in village, or district size human populations

The original models were designed for simulating the impact of interventions acting on humans(for example, chemotherapy, or vaccines). OpenMalaria has subsequently been extended to include simulation of the dynamics of malaria in the mosquito vector, and of interventions that act on mosquitoes. It also offers alternative model structures for many of the components of the human models. The models allow for variations among humans in their exposure to mosquitoes, and in their responses to the parasite.

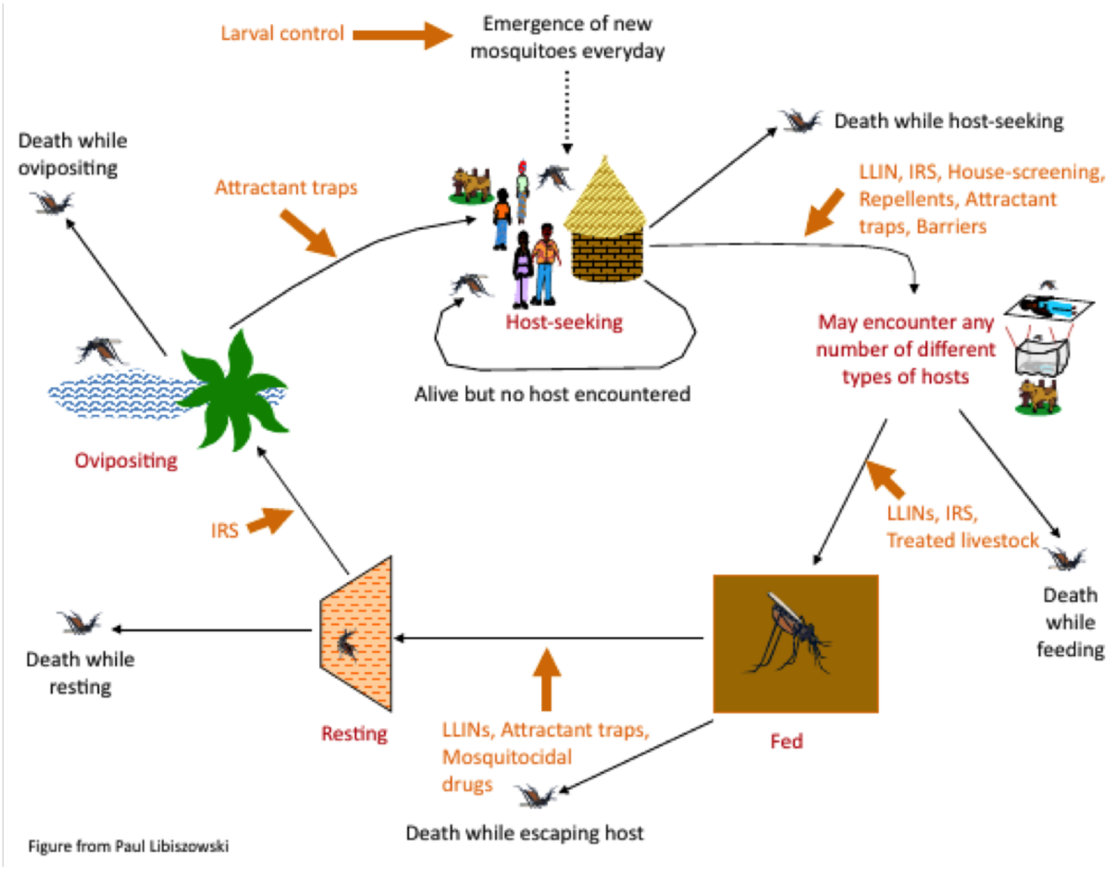
2 Facts about Malaria

- Endemic in 99 countries, 219 million cases, 660,000 deaths
- 90% in Africa
- Mortality rates fell by 26% around the world in the past decade

2.1 How does Malaria spread?

- Deterministic model of malaria transmission based on the mosquito feeding cycle
- Linked to models of interventions
- Driven by EIR

- For input EIR values, you may need to work backwards to determine pre-experiment level



Larval control
Emergence of new mosquitoes everyday
Death while host-seeking
Death while ovipositing
Attractant traps
LLIN, IRS, House-screening, Repellents, Attractant traps, Barriers
Host-seeking
May encounter any number of different types of hosts
Alive but no host encountered
Ovipositing
LLINs, IRS, Treated livestock
IRS
Death while feeding
Death while resting
Resting
LLINs, Attractant traps, Mosquitocidal drugs
Fed
Death while escaping host

Figure from Paul Libiszowski

Figure 1: The mosquito feeding cycle. Malaria is spread through this cycle.

3 How can Open malaria help in the real world?

It helps by enabling researchers to propose experiments, based on a number of observations that are collected on the field.

4 Simulation Options

The *OpenMalaria*simulator is designed for simulating individual scenarios. In order to simulate a study covering variations in several factors, one would have to design a fully *factorial* experiment.

- ***Sweep*** : A sweep is one of more covarying factors, and set of all *arms* of this sweep.

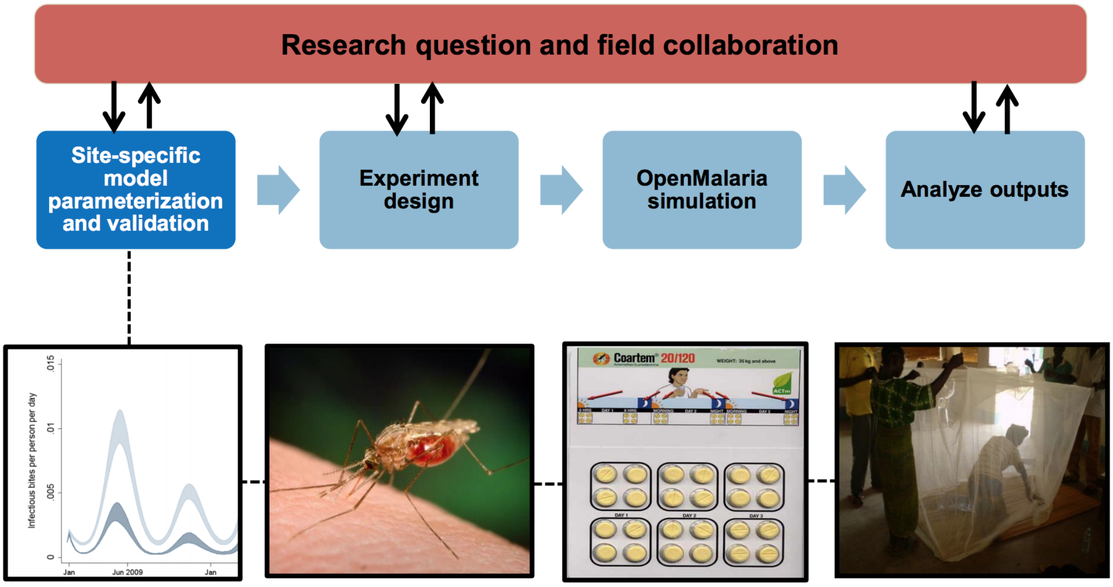

Research question and field collaboration
Site-specific model parameterization and validation
Experiment design
OpenMalaria simulation
Analyze outputs
Infectious bites per person per day
0
.005
.01
.015
Jan
Jun 2009
Jan
Coartem 20/120

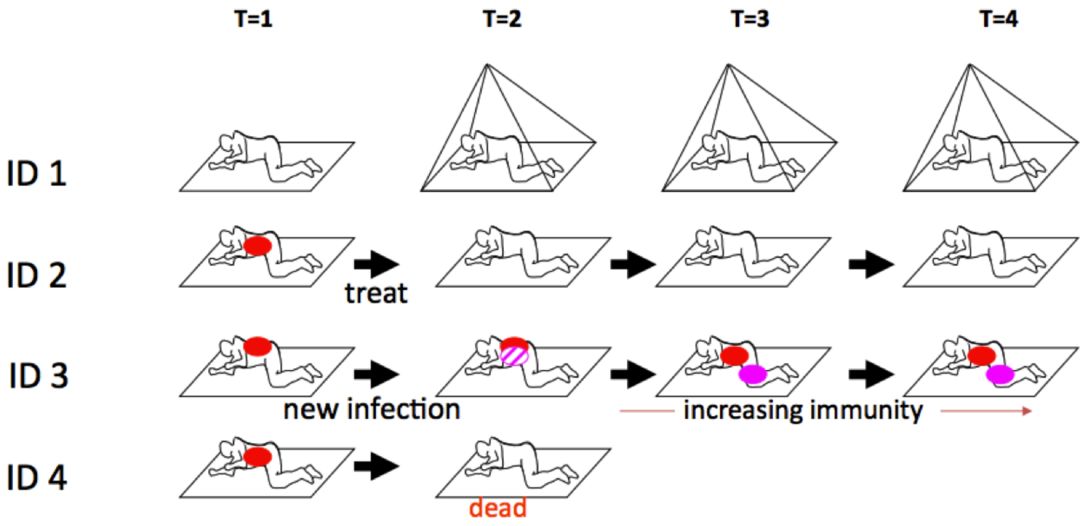
Figure 2: Where OM fits into the real world

- ***Arm***: Each sweep must have one or more values for each of its factors. An *arm* is one of these combinations of values - a configuration assigning a value to each factor on the sweep.
- ***Experiment***: Largely synonymous with a study, an *experiment* is the complete description of all sweeps and arms used, together with the scenarios generated and results produced. Each scenario is generated by choosing one arm for each sweep in the experiment.
- ***Full factorial design*** : A *full factorial experiment* is one where all possible scenarios are generated: all combinations of one arm per sweep are used.
- ***Seed***: Each time a scenario is simulated, the random number generated must be started from some *seed*.

Since OpenMalaria simulations are stochastic, they use a number of random seeds (say 2-50) to provide estimates of the contributions of random variation to the results.

5 XML files

Within the open malaria simulator *XML* files should be seen as the program, specifying a model over all parameters. Well kind of, it should actually be treated as a prior to the simulator, providing knowledge regarding how a model is specified in terms of the initial parameters. This then interacts directly with the simulator.



6 Key features of the simulator

- Seasonality of transmission
- Access to treatment an inherent part of model
- Addressing uncertainty: ensemble of 14 model variants

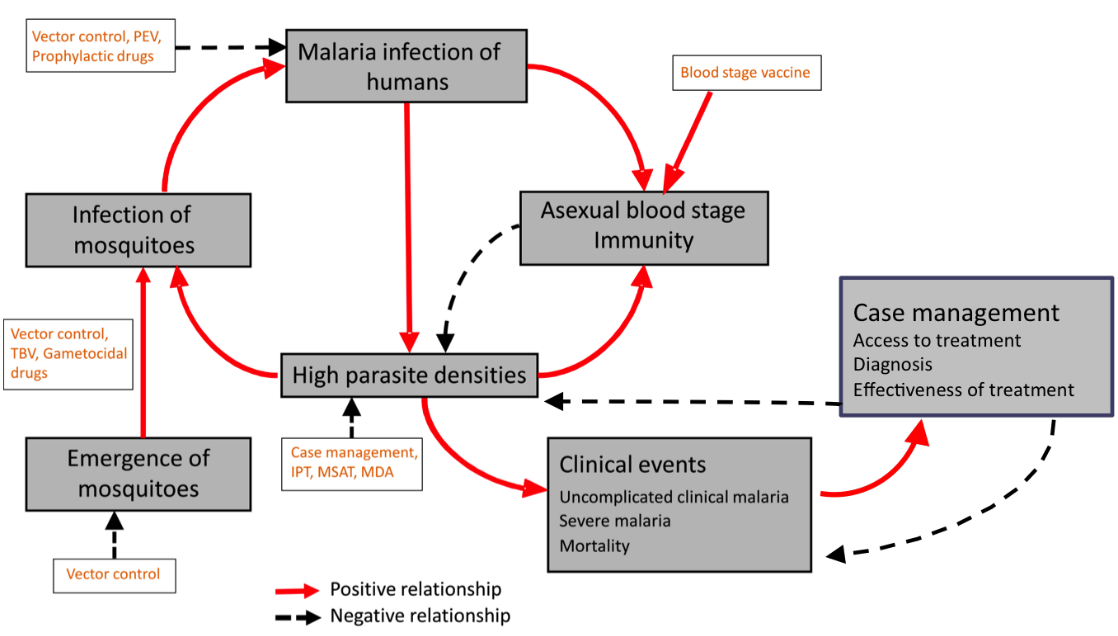
6.1 Modelling different treatments

Stochastic simulation models of malaria based on the simulation of infections in humans. Tracks health status of individuals at discrete time points

7 How to Use Open Malaria

Objectives

- Understand OpenMalaria workflow to design an experiment
- Understand the important features of OpenMalaria
- Understand, the topic of choice, the model assumptions and how interventions interact
- Understand what information is needed to parameterize OpenMalaria and where to find it



7.1 Seasonality of transmissions

The Plasmodium falciparum entomological inoculation rate (Pf EIR) is a measure of exposure to infectious mosquitoes. When describing transmissions, the simulator is driven by the input value of annual average EIR, the average number of infectious bites a person receives each year. The seasonality is specifically specified by 12 monthly values for each vector species. Abs values or proportion for each month.

In epidemiology, a **disease vector** is any agent that carries and transmits an infectious pathogen into another living organism; most agents regarded as vectors are organisms, such as intermediate parasites or microbes, but it could be an inanimate medium of infection such as dust particles.

7.2 Scenario transmission

OpenMalaria simulations of malaria transmissions require a specification of:

- The level and seasonality of exposure (measured by the Entomological Inoculation Rate, EIR) to malaria at the start of the simulation
- The model for transmission from mosquitoes to humans and the dynamics of malaria parasite cycle within humans.
- The model for malaria transmission from humans to
- The entomological model. There are two different variants of the entomological model:

- ⋆ The ”Non-vector” variant does not consider mosquito dynamics and hence uses fixed or periodically (seasonally) varying vectorial capacity. It is appropriate for modeling situations with only interventions such as chemotherapy or vaccines that only act on humans.
- ⋆ The ”Vector” variant comprises discrete-time population models that simulate how many mosquitoes belong in each of several categories at each time. The models assume that the infectious (sporozoite positive) mosquitoes act to distribute infections at random to the human population (with human exposure proportionate to availability). Entomological interventions modify the vectorial capacity and require the ”vector” transmission model variant. The simulations that include non-periodic changes in the vectorial capacity use a seasonally forced version of the difference equation model for vector dynamics.

7.3 Running a simulation

Within the same directory structure we add the following files:

- densities.csv
- scenario_xx.xsd

Then there must exist at least on file ¡MyScenario.xml¿

Then to run we use the following command: ./openMalaria –scenario MyScenario.xml

To ensure that it is possible to execute the openMalaria object, it must be within the execution path.

7.4 The phases of the simulator

- the XML document is read, model constants are set and some variables initialized
- an initial human population is created
- a warm-up period of one human life span is run
- there may be some extra simulations to adjust the mosquito population to the pre-specified Transmission situation.
- the intervention phase starts.

During these different phases the feedback from infections in humans to those in mosquitoes is handled in different ways

8 Gathering Data

The next step is to gather the necessary data to include in the scenario file to describe the situation in your study area as closely as possible to reality. There are more or less 90 separate pieces of data that need to go into the model. As many inputs as possible should come from studies in your particular study area, but realistically the overall percentage of inputs that come from site-specific data will be rather modest (about 30%). A sizeable number of inputs will come from previously published literature at the district, national or regional level through a literature review. There will be a number of inputs that will remain the same as the initial model parameterization, depending on which model(s) you choose, particularly data on the vector life cycle and drug action. It is helpful to document the source of the data you choose for each parameter and if there are multiple sources why you chose the value that will go into the model.

9 Outputs

The simulation results of the outputs you have chosen from Step 2 are produced in the form of two separate .txt files (ctsout.txt and output.txt) that can be imported into your program of choice for analysis (i.e. Excel, R, STATA, etc)

10 Tom stuff

Tom notes

- Always write down an abstract and introduction when you are starting a new project.
- Write down ideas as formally as possible.
- Things that happen first - the less applied part:
    - ⋆ As you mathematically formulate in code, you write down the mathematics that describes that and the problem formulation.
    - ⋆ Try and write down a joint of the model. That way you can determine if the problem is doable. Must deal with these things first.

Things to look into:

- Are there Multi-modes?
- Is the problem high-dimensional?
- What is the dimensionality of problems?

10.1 Approximate Bayesian Computation

When dealing with a constrained set of distributions we can perform exact Bayesian inference, that is we can use statistically correct inference algorithms to find the true posterior $p(x|D) = \frac{p(D|x)p(x)}{p(D)}$, not just heuristic results. If we choose the a conjugate-pair of prior $p(x)$ and likelihood function $p(x|D)$ then we can also analytically exactly calculate the true posterior. However, in general most real world problems, i.e. physically simulations, cannot necessarily be constrained to this family of problems, as the likelihood is typically intractable, or computationally expensive to compute. This occurs because the *forward simulator* emulates the likelihood function due to the ingrained stochasticity with the model, which makes it impossible to write down a likelihood function. Nonetheless, we can still apply a series of approximation techniques, such as variational inference that enable one to approximate the true posterior. Although such algorithms typically have no guarantees, they are typically lower bounded.

10.2 Applied Interpretable Inference

11 Gunes Notes

- ABC - talk about this in the context to MCMC
- The forward simulator represents some likelihood function that can be tractable or intractable.
- Within forward simulators the randomness comes from rejection sampling statements , but conditioning within the simulators is hard this is why we need to use the prob prog system .