Review: Modelling the transmission of Onchocerciasis

Karan Ruparell

University of Oxford

0.1 Acknowledgements

In this paper we cite private correspondence with a number of experts who were gracious enough to answer our questions. In particular, we would like to thank Professor Maria Gloria Basanez (Professor of Neglected Tropical Diseases in Imperial College), Dr Celia Petty (Strategic and Livelihoods lead at the Walker Institute), Dr Richard Selby (Head of Portfolio and NTD Research at SightSavers), and Professor Deirdre Hollingsworth (Leader of the neglected tropical diseases modelling consortium).

0.2 Abstract

Onchocerciasis, also known as river blindness, is a neglected tropical disease common in West Africa and parts of Latin America. Since its discovery in 1965, there have been many attempts to control the parasitic infection, and different models have been used in this effort. This paper outlines the different models used, providing an evaluation of how these have changed since their development and the increasing role of data analysis since 2016. We also discuss data limitations and the different model needs pre and post endemicity. We provide the first report to discuss all three models widely used in Onchocercaisis in a single paper, and the first explanation

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Onchocerciasis is a parasitic disease that effects millions of people, particularly those in poverty that cannot avoid areas of greater risk or easily be treated [1]. It causes great pain, and the global health charity Sightsavers estimates that 21 million people are currently infect with Onchocerciasis, 1.1 million of which are blinded by it. This gives it the name 'river blindness' [2]. The disease not only affects the lives of many but has the potential to affect even more. It is currently classified as a Neglected Tropical Disease by the World Health Organisation (WHO). This means it belongs to a group of 20 conditions that together affect more than 1 billion people has devastating consequences while being 'almost absent from the global health agenda'. 205 million people are at risk of contracting river blindness [3] [2]. Onchocerciasis leaves people blind, severely visually disabled, or in otherwise great pain. It can cause extreme itching, skin atrophy rashes, and nodules under the skin [2].

The World Health Organisation also speaks of the importance of cost-effective interventions. One avenue for achieving this is increasing our understanding of what areas are most conducive to the worm *Onchocercera volvulus*, so that mitigation efforts can be focused on these areas. Identifying suscpetible areas can involve incorporating domain specific knowledge, such as vector reproduction rates, into differential equation models. Alternatively, we can use data focused approaches to find high correlations of different features with infection rates. Two of the 9 research priorities for Onchocerciasis refer to the need for robust diagnostic tools and strategy optimisation, both of which can be helped by better modelling [3]. Better models allow for us to predict the impact of various control strategies on the prevalence and intensity of infection, and to evaluate the cost-effectiveness of different approaches in eradicating Onchocerciasis. In this paper we provide an overview of different models that have

been utilised, as well as their limitations and potential future work. In particular we discuss SEIS models, OnchoSim, EpiOncho, and Geosptatial Generalised Linear Models.

1.2 Background

Onchocerciasis is transmitted to humans by the bite of infected black flies that breed in fast-flowing waters. The black flies themselves are infected by the filarial worm Onchocera Volvulus, which lives in these fast-flowing waters, or are infected upon biting infected humans. Once they have been spread to humans they migrate to the skin, eyes, and other organs, where they cause great pain [2].

In 1974 the Onchocerciasis Control Programme (OCP) began in West Africa, which attempted to mitigate the spread of Onchocerciasis were done primarily by spraying insecticides against black-fly larvae. As the black-fly is the vector through which the Onchocera Volvula is transmitted to humans this approach mitigates the total number of infections by minimising the number of bites from the black-flies to humans. Treatment is long and potentially expensive, taking 12-15 years of annual treatment with Ivermectin [1] [4].

From 1989 this was complimented by the distribution of Ivermectin, a treatment for those already infected with Onchocerciasis. WHO recommends taking it at least once a year. It has been shown to reduce optic atrophy and improve lesions. Repeated community wide treatment can reduce transmission by reducing the odds of a single black-fly bite resulting in the black-fly ingesting the filarial worm. Since then Ivermectin has been the primary form of treatment, through Mass Drug Administration (MDA). From 1995 to 2015 this occurred under the African Programme for Onchocerciasis Control (APOC), was replaced in 2016 by the Expanded Special Project for Elimiation of Neglected Tropical Diseases (ESPEN), this is planned to run until 2025 [5] [6].

1.3 Functions and Notation

Exponential Distribution: for $x \in (0, \text{inf})$ an exponential distribution with mean λ has distribution

$$f(x) = \lambda e^{-\lambda x} \tag{1.1}$$

Poisson Distribution: for $k \in \mathbb{N}_0$

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{1.2}$$

Weibull Distribution: for $x \in (0, \text{inf})$ a Weibull distribution with scale λ and shape k has distribution

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \tag{1.3}$$

Gamma Distribution: for $x \in (0, \text{inf})$ the Gamma distribution is defined as

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{1.4}$$

Generalised Linear Models (GLMs): A generalised linear model is a model in which the hyper-parameters of the distribution for some output Y is defined to be a function of a linear combination of the feature variables X, i.e.:

$$p(Y = y|X) = f(y|\theta), \quad \theta = h(b + \sum_i \alpha_i \times X_i) \tag{1.5}$$

For a binomial GLM predicting the number of cases m in a population of size, n under the canonical link function, the function is as follows:

$$p(Y = m|X) = (mCn) \cdot \pi^m) \cdot (1 - \pi)^{n-m} \tag{1.6}$$

$$ln(\frac{\pi}{1-\pi}) = \alpha + \beta X \tag{1.7}$$

Stochastic Processes: A stochastic process is a family of distributions, usually with some index. A two dimensional index for a stochastic process may then be written as

$$(S((x_i, x_j))_{(x_i,x_j)\in X} \tag{1.8}$$

Where there is often some correlation between the the distributions at different indices.

Semivariogram: A semivariogram $\gamma(d)$ with respect to a function f describes half the average squared difference of the outputs between two points a distance d away . For a given distance metric D this is written as

$$\gamma(d) = \frac{1}{2V} \int_V \int_{v':D(v,v')\in V} [f(v') - f(v)]^2 dV' dV \tag{1.9}$$

The semivariogram requires all combinations of points a distance d away to be sample, and so is in practice intractable. The empirical semivariogram is thus calculated as

$$\hat{\gamma}(d) = \frac{1}{2|N(d \pm \delta)|} \sum_{d(i,j) \in N(d \pm \delta)} |z_i - z_j|^2 \tag{1.10}$$

For locations i and j with corresponding observations z_i and z_j. Often these empirical semivariograms are fit to a parametric model, and for the purposes of this paper we will show the exponential correlation function

$$\rho(d) = exp(-d/\phi) \tag{1.11}$$

Lifetables

Lifetables show the probability of a randomly selected individual in a population surviving to the selected age. The survival probability at an age a is $F(a)$

Example:

age (a)	0	5	10	15	20	30	50	90
$F(a)$	1.000	09804	0.872	0.750	0.640	0.581	0.309	0.000

Chapter 2

The use, or lack thereof, of SEIS models in Onchocerciasis

2.1 Model

The use of SEIS (Susceptible-Exposed-Infected-Susceptible Treated) models in disease modelling is well known as a result of the COVID19 virus, and is a very useful tool for modelling the spread of many diseases including Dengue fever [7] [8]. A generalisation of SIR (Susceptible-Infected-Recovered) models, SEIS have also been used in a number of papers to model Onchocerciasis, however these papers have been limited primarily to master's projects and are not used by the World Health Organisation and other bodies [9] [10] [11]. Through private correspondence with several academics leading the research efforts in Onchocerciasis, we are able to offer the first written section evaluating why SEIS models are believed to be unsuitable for modelling Onchocerciasis.

During our investigation into models used for modelling aspects of Onchocerciasis, we found a number of papers using SEIS models [12] [10] [9].

- Susceptible $H_s(t)$: Those that could be exposed to the disease
- Exposed $H_E(t)$: Those exposed to the disease but not yet infected
- Infected $H_I(t)$: Those exposed and infected
- Susceptible and treated $H_T(t)$: those susceptible but have undergone treatment

Sometimes, the exposed and infected groups are treated identically, resulting in a SIS model. In the case where treated and recovered people cannot get the infection again, this is a SIR model.

This framework does not usually incorporate the behaviour of vectors, however it can be easily adapted to do so. In 2014, Oguoma and Acho were the first to adapt these models for use in modelling Onchocerciasis introducing Vector susceptibility V_s and Vector infectious V_i groups [12]. The models consist of 3 partial differential equations:

$$\frac{dH_S}{dt} = \rho\zeta + \gamma V_I - (\beta_1 + \beta_2)H_S \tag{2.1}$$

$$\frac{dH_E}{dt} = \delta\lambda + \beta_1 H_S - \beta_3 H_E \tag{2.2}$$

$$\frac{dH_I}{dt} = \beta_2 H_S + \beta_3 H_E - \delta\lambda - \rho\zeta - \beta_4 H_I \tag{2.3}$$

$$\frac{dV_S}{dt} = \beta_4 H_I - \beta_5 V_S \tag{2.4}$$

$$\frac{dV_I}{dt} = \beta_5 V_S - \gamma V_I \tag{2.5}$$

There have been some other papers further applying SEIS models to Onchocerciasis, but they seem to be relatively unused and are not in reports by WHO and other model reviews around Onchocerciasis [9] [13] [11] [14].

2.2 Intensity based modelling and the problem with SEIS models

SEIS models assign individuals to one of 4 categories, and do not distinguish between levels of illness beyond 'Susceptible' and 'Infected'. This is a prevalence based model, as it models only whether or not a disease is present within an individual, and not the strength of that infection. Prevalence based models are useful for diseases such as Covid19 where the disease grows quickly inside the body when someone is infected. This makes the variable of most concern the prevalence of the disease, as once inside an individual there is relatively little work to be done in minimising the intensity inside a person, beyond single dose treatment which is represented in the model. This is not true in the case of Onchocerciasis however, where Ivermectin treatment is used to decrease the density of MF within the body. The density of MF plays an important role in different dynamics of Onchocerciasis, including transmission dynamics, morbidity, population regulation processes. The average density of an infection per individual is a definition of intensity, and gives rise to intensity based models[15]. If prevalence had a 1-1, or near 1-1, relationship with intensity then modelling intensity would also be a suitable method as it would inherit the same relationship with

transmission dynamics and other processes. As Remme et al. showed in 1986, however, this is not true [16]. Instead, different prevalence values can even correspond to different exposure heterogeneity and blackfly to human ratios [17].

Through private correspondence with Professor Maria Gloria Basanez, a coauthor in many of the papers in the field including the original EpiOncho paper, we believe that the reason SIR models are less relevant in modelling Onchocerciasis is that it is a prevalence based model [18] [17] [11]. . Prevalence based models focus on the fraction of the population that has been infected, intensity based models describe the average number of parasites present in a host [15]. In contrast, the level of intensity of *Onchocercera volvula* does play an important role in different dynamics of the disease, and as Remme et. al have shown in 1985 this is not linearly correlated with prevalence [16]. Hamley et al. write that the same prevalence values can correspond to different combinations of exposure heterogeneity and black-fly to human ratios, and advise against the use of prevalence alone to determine intervention success [17]. The intensity determines morbidity and whether or not someone will be blinded by the disease, and it also effects transmission dynamics . It is for this reason that SIR based models are unsuitable for modelling Onchocerciasis [19].

Prevalence models can be useful in mathematical models pertaining to Onchocerciasis, but they are unsuited for modelling the dynamics of the disease. In chapter 5 it will be shown how they have even been used successfully [20] [21].

Chapter 3

1990: OnchoSIM, computer simulations

3.1 History

The first computer program for modelling Onchocerciasis, OnchoSim was developed by Plaiser et al. in 1990 to model and control the spread of Onchocerciasis [22]. A number of organisations were involved in the making of this model, including the Onchocerciasis Control Programme. This means that the authors were likely in direct contact with the people working on implementing policies and treatment efforts for Onchocerciasis, and so were able to design a need focused approach to the modelling problem.

3.2 Model

OnchoSim consists of 35 parameters, covering aspects such as human population dynamics, larval population dynamics, infection rates, and the effect of treatment. The full list can be found in the original paperp [22]

The model works by splitting the simulation into separate systems that interact with each other. These are

1. Human population demography: Number of men and women, survival probability as a function of age, number of offspring per year as a function of age
2. Exposure: p.d.f of individual exposure risk, conditional on age and sex E_{xa}, and an individual exposure index E_{xi}
3. The parasites' behaviour inside a human host: lifespan, prepatent period, microfilarial productivity etc.

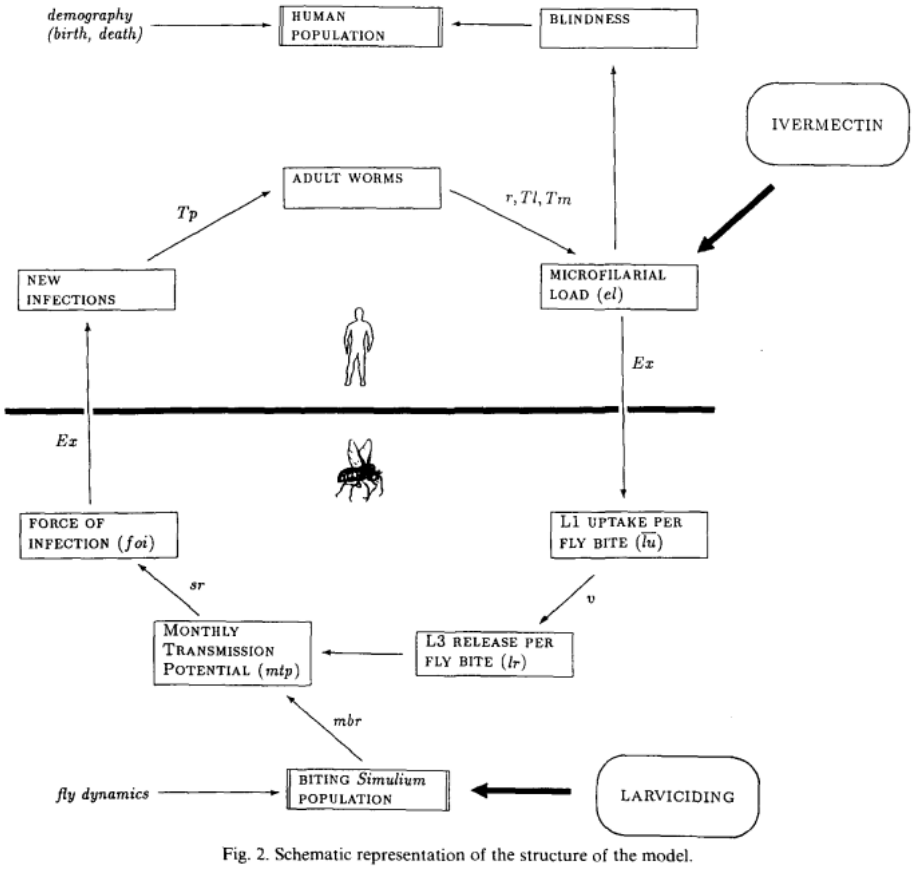

Fig. 2. Schematic representation of the structure of the model.

Figure 3.1: OnchoSIM dynamic model

4. Fly and larval dynamics

5. Disease behaviour - excess mortality, blindness threshold

6. Ivermectin - effect and coverage

7. Larviciding - inter-application period and coverage

This method can be summarised in the figure below:

While a full description of each of the parameters is unwarranted for this project, we provide a brief overview of the key aspects of the model.

3.2.1 Individual Exposure

Noting that age and sex impact an individuals risk of becoming infected, as well as other factors that aren't directly observed, OnchoSIM gives each individual x_i in the model an independent exposure level Ex_i, which is comprised of an age and gender

specific risk and an inherent stochastic risk. The exposure Ex_i of an individual x_i is calculated as

$$Ex_i = \text{Exa}\,(a_i, s_i) \times Exi \tag{3.1}$$

Here $\text{Exa}\,(a_i, s_i)$ is a deterministic risk dependent on an individuals age and sex, and Exi is generated from a random probability distribution in the range $0, \text{inf}$. Often, the distribution of Exi is taken as $Exi \sim Gamma(1, \alpha_{Exi})$.

The number of bites $mbr_i(t)$ a person gets in month t is then calculated as

$$mbr_i(t) = Mbr(t)Ex_i \tag{3.2}$$

where $Mbr(t)$ is the number of bites a person of relative risk 1 received. This is a set of 12 parameters chosen to align with data collected over 6 years from Absande (Ghana) [22].

3.2.2 Connection to observed data

To validate this model it is important to have an element that can be compared with real data. As such, one of the most important parts of the model is its prediction on the density of microfilariae in the skin, as this is what is checked in mass testing for Onchocerciasis. This is important because, the density of worms in a host is the most crucial factor for it's effect on people, and not merely it's prevalence.

The expected number of skin snips per 2mg is given as:

$$ss(t) = \frac{cw}{Tm} \sum_{j=1}^{n_i} d_j \sum_{x=1}^{Tm} r_j\,(a_j - x, t - x)$$

Where d_j is the dispersal factor for female parasite j, which is randomly drawn from an $Exp(1)$ distribution for every worm at birth.

Individuals are modelled as going blind when their cumulative parasitic load goes beyond a Weibull distributed individual threshold level.

3.2.3 Stochastic Behaviour

As this model is a stochastic model, there are multiple actions that can occur at every time step. At each time step, a worm may mature, mate, or die. A human may

1. Become infected with new parasites
2. Become blind

3. Die

4. Have a skin snip taken from them

5. Be treated with a microfillaericide

In particular, 4 can be used to validate our models.

Plaiser et al. [22] uses this model to simulate the situation in Folonzo, Burkina Faso, before and after Onchocerciasis control. They model the distribution of microfilariae per skin snip in individuals in Folonzo before and after control, and find that the the model has some success in doing this. They only look at populations of size 30-40 however, and do not attempt to quantify the difference in distribution between both distributions.

3.3 Work that followed

OnchoSim was also used by the OCP [23]. In 1999, the Expert Advisory Committee for the OCP released a report evaluating existing OCP efforts using OnchoSim and generating recommendations for future work as a result of simulations from the model.

In that paper their work was limited to unrealistic assumptions on the effectiveness of Ivermectin and control efforts, assuming that control completely stopped transmission with the considered area. They produce simulations of Onchocerciasis transmission within an area, comparing the time needed for the elimination of the parasite, depending on what control measures were used, concluding that while Ivermectin can help mitigate the spread of the parasite, vector control is much more important. They advise strong caution with the use of these models however, calling primarily for more data, more multidisciplinary analysis of data, and cautioning against making strong conclusion on single simulations. Multiple trials were not run, and this appears to be because of computational limitations. In 1999 the strongest computers available had roughly 32MB of DDR memory, compared to 8GB in most laptops today. More recently, OnchoSIM was used in 2014 to model the impact of increasing Ivermectin treatment frequency, and has also been used to evaluate the effectiveness of APOC, estimating the saving of 17.4 million DALYs (Disability Adjusted Life Years) [24] [25].

Chapter 4

1999: EpiOncho

4.1 Model

Unlike OnchoSim, the EpiOncho model is deterministic. It was first introduced in 1999 by Basanez and Boussineq, and uses partial differential equations to model the relationships between worms, flies, and [18]. This first model did not include an explicit age structure, and it also assumes that the density of humans is a constant H in time. Similarly, the density of vectors is also a constant V. As a result, all the equations that determine the behaviour of the model refer to a single individual, and the relative number of worms and flies per individual. This model takes the form:

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = \frac{V}{H}\beta\left(\frac{\delta_{H_0} + \delta_{H_\infty} c_{H_i}(V/H)\beta L(t)}{1 + c_{H_i}(V/H)\beta L(t)}\right) L(t) - (\sigma_W + \mu_H)\, W(t) \tag{4.1}$$

$$\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \left(\frac{1}{2}\phi F\right) W(t) - (\sigma_M + \mu_H)\, M(t) \tag{4.2}$$

$$\frac{\mathrm{d}L(t)}{\mathrm{d}t} = \beta\left(\frac{\delta_{V_0}}{1 + c_V M(t)}\right) M(t) - \left(\sigma_L + \mu_V + \alpha_V M(t) + \frac{a_H}{g}\right) L(t). \tag{4.3}$$

Here, $W(t)$ is the number of adult worms per person at time t, $M(t)$ is the number of of microfilarieae per milligram of skin per person, and $L(t)$ is the mean number of L3 larvae per fly. While there are too many terms to define in this review, a full list of the parameters and their definitions can be found on page 7 of the original paper [18]. The model parameters fall under 5 main categories:

- Host and Parasite populations V (vector) , H (Human hosts), W (worms), M (MF per milligram), L (infective larvae)

- Demographic rates: per capita death $\mu_H, \mu_V, \sigma_W, \sigma_M, \sigma_L$, fecundity F, mating probability ϕ
- Transmission rates: biting rate per human β, proportion of L3 larve shed per bite α_H, proportion of L3 larvae developing into adult worms in the human δ_H, human-blood index h etc.
- Regulation of parasite population abundance: defining the limit of different processes as terms tends to 0 and ∞, marked by parameters with subscript 0 or inf
- Behaviour of larvae within the vector host: proportion of microfilariae developing to infective stage in the host per bite, excess vector mortality, development of larvae

By solving these equations the paper is able to find a formula for R_0, the reproductive rate at which the density of microfilariae in humans is stable, and below which endemicity cannot be maintained. This is important in order to have an idea of how to prevent and stop endemicity of Onchocerciasis in different areas, and assess how effective different measures can be in limiting it. The R_0 found is shown below

$$R_0 = \frac{\phi F(V/H)\beta^2 \delta_{H_0} \delta_{V_0}}{2\left(\sigma_W + \mu_H\right)\left(\sigma_M + \mu_H\right)\left(\sigma_L + \mu_V + (a_H/g)\right)} \tag{4.4}$$

4.2 Work that followed

There are two key concerns with this first specification of the EpiOncho model. Firstly, as we have discussed in the Criterion section, this model is deterministic. The model also makes an assumption that each human has an equal level of risk to the disease. This is untrue, as is pointed out in the paper itself. While the model is often in agreement with other models and results, sex and age do play important roles in rates of infection and so it would be useful for it be incorporated into the data.

The EpiOncho model has since been developed, and the latest deterministic version was developed in 2017 by Walker et. al [26]. This model splits the larvae population into its different stages of development, $L1, L2$, and $L3$. It also splits humans into different demographic group, in order to account for age and sex. It replaces various parts of the model with more realistic mechanisms, to better incorporate the increased prevalence of MF after the initial rounds of Ivermectin MDA and better

predictions of parasite survival times among other features. In this paper parameter sets were determined using a Sampling Importance Resampling approach, which requires the specification of a prior. This prior was then also used to find the Maximum a posteriori parameter set, as an alternative method of defining the parameter sets.

After this, Hamley et. al incorporated a stochastic aspect into EpiOncho, resulting in the model EpiOncho-IBM [17]. In 2019, Hamley et al. attempted to recognise these differences by adapting an EpiOncho model to give individuals i a randomly assigned exposure factor, $E_{(i)}$. This is randomly distributed under $E_{(i)} \sim gamma(k_E, \beta_E)$. The density of MF in each human in a population is then simulated under it's own EpiOncho model.

Chapter 5

2016: Bayesian model-based Geostatistical mappings

In 2016 O'Hanlon et al. developed a statistical model using geospatial methods in a binomial GLM. This seems to be the first model of Onchocerciasis that doesn't use a mechanistic understanding of how Onchocerciasis is transmitted, instead focusing on fitting observations onto a statistical model. The model operates as a binomial GLM with the canonical link function, as introduced in the background section of this essay [20]. One key determinant for how feasible Onchocerciasis elimination is in a given area is the pre-control level of endemicity in that area, an indicator for how widespread the disease would be in the given area if no actions were taken. Endemicity levels are often defined by the Community Microfilarial Load (CFML), the geometric mean number of microfilariae per skin snip in adults.

5.1 Method

Briefly, the model is tasked to predict the number of found cases of Onchocerciasis Y_i in implementation unit i out of the total number of tested people N_i, given information about location i in the form of row X_i. The full model can be expressed as $Y_i \sim B(N_i, \pi(X_i))$. They also include an 'unobserved, underlying spatial process', designed to incorporate similarities between regions that are close together that are not related to other recorded features. This is modelled by a stochastic process $S(X_i)$ with expectation 0, variance σ^2, and covariance function,

$$C(d) = \left\{ \begin{array}{ll} \sigma^2 + \tau^2 & : d = 0 \\ \sigma^2 \rho(d) & : d > 0 \end{array} \right\} \tag{5.1}$$

and $\rho(d)$ the exponential correlation function $exp(-d/\phi)$. This stochastic process, $S(X_i)$, provides a non deterministic model for the correlation between location not

explained by the variables used in the model. Samples are generated from it using Markov Chain Monte Carlo simulations to sample the value of the stochastic process at a new location, $[S(x^*)|Y, \beta, \theta]$, as it is not possible to directly draw from this distribution. The stochastic process is important as it is almost never possible to have all factors important for determining the spread of Onchocerciasis in a model, both because of the impossibility of knowing which features are relevant and the intractability of collecting such data. However, it can often be assumed that many features that are relevant but unknown are similar for locations that are similar. For example, it may be reasonable to assume that, on average, cities that are close together may have similar wealth and primarily livelihoods for individuals, and so are correlated through this. We also assume that far apart locations are less likely to be similar, and so are less correlated. The exponential correlation function incorporates this into the stochastic process, so the value at locations close together are more correlated than those far apart. The loss of correlation, how far two locations need to be before they are roughly uncorrelated, is controlled by the 'sill' or 'dispersion' parameter.

5.2 Work that followed

After this original paper in 2016, Retkute et al. furthered this work with some of the original authors in 2020. They use Adaptive Multiple Importance Sampling (AMIS), a deterministic multiple mixture model that is less subject to the curse of dimensionality as other sampling techniques [27]. At each iteration samples are removed that are over represented in the sample, as measured by Kish's effective sample size. This makes the model more robust, generating proposals more suitable for underrepresented samples. This is especially important when there is larger uncertainty in measurements and an absence of data, as in the case of Onchocerciasis prevalence as noted by O'Hanlon et al. [20]. In line with ESPEN's movement towards working towards the mitigation of NTDs through collaborative research across them, Retkute. In the case of Onchocerciasis, the goal of this paper was to make projections for the spread of the disease using the latest variant of the deterministic EpiOncho model and data from the initial paper by Hamley et al.. They use a uniform prior for the log of annual biting rate and the aggregation of adult worms, decisions which are not justified in the paper. They simulate the effect of 65% and 80% coverage of Ivermectin treatment, corresponding to the minimum recommended coverage as per the World Health Organisation and an enhanced level of coverage respectively.

The results show that while both strategies appear to have very similar mean prevalence path over the course of a 15 year period post treatment at all three levels of endemicity, the uncertainty is greatly decreased for 80% coverage.

In 2021, a large project, including over 100 authors and 200 named institutions, produced a paper statistically evaluating environmental suitability to Onchocerciasis in over 2400 implementation units (IUs). This was a a large data collection effort: the majority of reports were from national bodies; some data had to be requested from the former director of the OCP; and more geographical data had to be extracted and standardised [21]. Professor Basanez, coauthor in both the 2016 paper by O'Hanlon et al. and the 2021 paper by Cromwell et al., has revealed through private correspondence that further work is being carried out current to combine the climatic data analysis performed here with a stochastic variant of EpiOncho described previously, in order to model host heterogeneity and Annual Biting Rate (ABR) [19]. This will be part of a project funded by the Bill and Melinda Gates foundation.

The use of statistical models for modelling Onchocerciasis has become increasingly popular since the 2016 paper of Hamley et al.. As the field is neglected, there are relatively few researchers in the field, and so the lead authors of the 2016 paper have since been authors in much of the other papers following this work, however it has been used in papers where none of the original authors were present [28]. These statistical techniques have been used primarily with the aim of prediction, finding hot spots where endemicity is particularly likely for environmental reasons. They have also been used in conjunction with deterministic models, using them to efficiently predict a suitable and robust joint distribution of climatic and model parameters that represent a diversity of location types, and using these models to simulate the outcome of interventions.

Chapter 6

General Limitations

Mathematical models are often constrained by two key factors: access to data and computational budget. This can also be seen in the history of modelling in Onchocerciasis. Here, we discuss these constraints and the role they have played in model usage and development in Onchocerciasis.

6.1 Data access

Data analysis methods for modelling the spread of Onchocerciasis aim to predict the number of cases of Onchocerciasis in given regions. This means that they rely on training data which contains the correct, or close to correct, estimate of the number of cases in different regions. Even for dynamic methods, however, data access is important for the tuning of parameters in the model, and so the need for high quality data has existed since the start of the Onchocerciasis Control Progamme (OCP).

6.1.1 International data gaps

One form of data gap that exists and is important in the modelling of Onchocerciasis is that levels of data access vary significantly between countries. In particular, many places with high levels of endemicity also have very little data, such as the Democratic Republic of Congo (DRC) [29]. We held an interview with Dr Richard Selby - Head of Portfolio and NTD Research at the non-profit SightSavers - who explained this further [30]. The DRC has little travel infrastructure in comparison to other countries where Onchocerciasis is common such as Nigeria. With a strong background in field work, Selby talks of a single tarmac road in the countries, away from which they were only able to travel at 20 miles per hour. A lack of accessibility greatly increases the cost of collecting data from different areas, and, as noted in our introduction, the field

of Neglected Tropical Diseases is heavily funding constrained. Selby also spoke of conflict and instability as a limiting data access. The DRC has a large amount of natural resources, and there are often conflicts of land ownership and instability as a result [31]. In order to collect data on the number of cases in different areas people need to go there to collect the data, and this is not feasible in many cases. A desire to be able to focus resources in areas where endemicity is particularly likely, even when mass testing has not yet taken place there, seems to be a key motivation in some of the climate focused research that organisations are doing around Onchocerciasis. Sightsavers and GLIDE are working on creating models that can better predict which fast flowing rivers are likely breeding sites for the blackfly that transmits *Onchocercera volvulus*, and the use of geospatial logistic regression by Cromwell et al. was designed to uncover which of 2400 untested Implementation Units were most likely to be at risk of endemicity. Selby says these studies are particularly useful in telling us not only which places are most at risk, but also what places are likely not to be at risk [30].

6.1.2 Within region data gaps

The data we do have however is not very representative. O'Hanlon et al. point out that although many surveys have been conducted in different places, each one was only able to sample a small fraction of the population [20]. Moreover, these samples are not representative. This is because surveys are often conducted in areas where we know there is great risk, creating bias towards areas of high endemicity. There is also a bias towards people who have the means to go to clinics and get tested, which is not viable for those who are seriously ill or poorer and so have less time or access to transportation.

In an interview, Dr Celia Petty also discussed the potential importance of different demographic factors [32]. Dr Celia Petty is the Strategic and Livelihoods lead at the Walker Institute and has worked extensively in international policy, including work on Onchocerciasis. Citing work from Ogebe et al., Petty points out that while the cost of an individual being burdened with Onchocerciasis is great, the cost of going to receive treatment may also be great [4]. Ogebe et al. estimate the cost of Onchocerciasis prevent to be approximately 34,000 Naira per year (£59). The cost of illness per household per year they estimate as 253,000 Naira (£440). While this difference is great, it is unclear if the cost of treatment is affordable to all individuals. Ibe et al. estimated that the cost of the drug administration was roughly 6× as much as a caregivers monthly income [33], and that many with Onchocerciasis did not know the

cause of their symptoms. These suggest that economic factors may play a key role in deciding who does and does not get treatment in a community, and so who is mostly likely to face greater burden of disease. Petty believes that more research is needed in order to better determine whether or not wealth is a critical factor that needs to be incorporated into models, but as of yet there is limited data on this. She says that while village studies are likely to give you information from a wide range of economic groups, this is often not done in favour of random samples as it is quite expensive to perform. When it is done, wealth disaggregation is uncommon. There may also be an intersectionality between wealth and gender, Petty says, as girls from poorer communities may be less likely to attend school where treatment is often done, or may have less freedom in travelling to clinics.

6.1.3 Collating the data

There is also great difficulty in collating data when it is available, as much of the data comes from different databases that need to be connected [27] [20]. The earliest model we studied, OnchoSIM, was only able to use a study surveying 612 individuals, from a single village. No data beyond MF density in each individual was collected.

The first large push towards using more data seems to have occurred in 2016 by O'Hanlon et al. They combined climatic data from a number of different sources, as well as a number of different resolutions. They covered a large amount of information including vegetation index, land surface temperature, land cover type, river flow direction and accumulation, and many others factors. This work made it easier for future work to incorporate the effect of climatic features of Onchocerciasis dynamics, and was likely also used in shifting norms towards data analysis [27]. Since this paper multiple papers have been produced utilising more data, including a large 2021 effort collating data from multiple national bodies [21] [28].

Despite this, data concerns still remain. In order to have a satisfactory amount of data to have strong predictive power of pre-intervention endemicity of areas, O'Hanlon et al. had to use data from between the years of 1974 and 1990, ignoring the effect of time on the variables. While the authors conclude that there seemed not to be a temporal trend, it is likely that one would emerge once we were able to incorporate more variables into the model, or had more data to increase our confidence in the effect of different variables. In 2021, data is collected from 1975 to 2017, an even larger range in order to make more specific conclusion in identifying at risk areas previously untested.

6.2 Communication and model needs

Selby describes one key difficulty in the use of models in Onchocerciasis to be in communication [30]. They describe conversations with Ivermectin treatment implementers that would benefit from the use of models such as EpiOncho for decision making, but were unaware of it. One reason for this, Selby believes, is that academia is often separated from policy, and maximising citations are not related to maximising usage. While Basanez and others work intensely with the WHO and the Bill and Melinda Gates foundation, and SightSavers invest a substantial amount in knowledge dissemination, there is still a large gap between research and local actors [21]. It is only after communication with Deirdre Hollingsworth that we became aware of the fact that PENDA, the original planned replacement for APOC, was superseded by ESPEN [34] [6]. Even with this knowledge, we were unable to find reference to this change online. This suggests that more communication and public dissemination knowledge is needed to encourage awareness of latest developments in Onchocerciasis.

The need for model use by local groups also limits the potential complexity of these models. In many countries where Neglected Tropical Diseases are rampant, such as Yemen, internet access is limited [35]. When internet is available, it is often unreliable. Combining this with the fact that many implementers rely on old laptop, Selby described it as a reasonable assumption that most people in the field work on laptops, there is a strong constraint to model complexity. OnchoSim and EpiOncho are able to be run multiple times to get increased confidence in results and elements can be added to increase their complexity. This is a major advantage they have over other models, such as machine learning techniques, as their most basic versions can be run even on very limited devices. This is because they were designed in the 1990's, where compute was far more limited than it is today. For cloud based solutions, unreliable internet access means data must ideally be able to be uploaded in a few minutes rather than relying on stable connection for the span of an hour or more. However, this is not to say that more complex models can't be used for large scale planning, such as for deciding which locations to prioritise in international efforts. These efforts do not require models to be run by individual policy makers, instead requiring only a body of individuals such as ESPEN. This means that larger models may still be useful in international cooperation, as in the paper by Cromwell et al. [21].

In the spirit of this, I asked Selby about the importance of social dynamics in modelling Onchocerciasis, as thus far models in Onchocerciasis have treated the rise

of Onchocerciasis in individuals as being independent conditional upon other local parameters such as fly density [30]. Their response was that while social dynamics may make the models more accurate,the current stage of the Onchocerciasis control is wide scale mitigation, and for this the models currently in existence are suitable. They believe that more sophistication in model dynamics may be useful, after the current phase of Onchocerciasis control. There are still communities that models fail to accurately simulate the behaviour for, either because Ivermectin coverage is sufficiently different or because social dynamics are substantially different. Nditancho et al. describe strong differences in semi nomadic groups in the Cameroon compared to nomadic groups, ranging from cultural and language barriers to mistrust. These differences have impacted the success of NTD interventions in the past, and will likely continue to do so. Selby claims that while this is an important problem the limited resources may be better focused on helping the majority of communities. They believe that as incorporating social dynamics inherently requires individualising models to different communities, it is unsuitable for using across regions without individualising them which is costly. However, many aspects of social models, such as that of triadic closure, hold throughout communities, and so this may be both generalisable and usable. It has also been used in other areas of disease control [36].

Chapter 7

Conclusion

In this paper we have provided an overview of models used in evaluating and forecasting the spread of Onchocerciasis and how these models have been developed since 1990. We have explained the present requirements for models in this field. We have used this to discuss why SIR models are unsuitable for modelling this disease and why current model improvements focus on the incorporation of data that is easily accessible across countries. We have also discussed the growing needs of these models, offering insights into how the models needed for mass implementation differ from the models needed for local eradication, and what effect that may have on future research. This research would not have been possible without the numerous experts that we were able to consult, and who have allowed us to incorporate the views of field workers, modellers, and policy makers within this report.

Bibliography

[1] World Health Organisation. Onchocerciasis, 2022.

[2] Sightsavers. What is river blindness?

[3] World Health Organisation. Neglected tropical diseases, 2023.

[4] FO Ogebe, GA Abu, and UC Nnama. Assessment of economic cost of river-blindness (onchocerciasis) illness among rural farming households of benue state, nigeria.

[5] World Health Organization and African Programme for Onchocerciasis Control. African programme for onchocerciasis control (apoc). Technical report, 1999.

[6] Adrian D Hopkins. Neglected tropical diseases in africa: a new paradigm. *International health*, 8(suppl_1):i28–i33, 2016.

[7] Ian Cooper, Argha Mondal, and Chris G Antonopoulos. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, 2020.

[8] Abhishek Pandey, Anuj Mubayi, and Jan Medlock. Comparing vector–host and sir models for dengue transmission. *Mathematical biosciences*, 246(2):252–259, 2013.

[9] EO Omondi, F Nyabadza, and RJ Smith? Modelling the impact of mass administration of ivermectin in the treatment of onchocerciasis (river blindness). *Cogent Mathematics & Statistics*, 5(1):1429700, 2018.

[10] Abdon Atangana and Rubayyi T Alqahtani. Modelling the spread of river blindness disease via the caputo fractional derivative and the beta-derivative. *Entropy*, 18(2):40, 2016.

[11] NTD Modelling Consortium Onchocerciasis Group. The world health organization 2030 goals for onchocerciasis: Insights and perspectives from mathematical modelling: NTD modelling consortium onchocerciasis group. *Gates Open Res.*, 3:1545, September 2019.

[12] Ikechukwu Chiwueze Oguoma and Thomas Mbah Acho. Mathematical modelling of the spread and control of onchocerciasis in tropical countries: Case study in nigeria. In *Abstract and Applied Analysis*, volume 2014. Hindawi, 2014.

[13] Asha Hassan and Nyimvua Shaban. Onchocerciasis dynamics: modelling the effects of treatment, education and vector control. *Journal of Biological Dynamics*, 14(1):245–268, 2020.

[14] María-Gloria Basáñez, Martin Walker, HC Turner, LE Coffeng, SJ de Vlas, and WA Stolk. River blindness: mathematical models for control and elimination. *Advances in parasitology*, 94:247–341, 2016.

[15] Allison K Shaw, Julie Sherman, F Keith Barker, and Marlene Zuk. Metrics matter: the effect of parasite richness, intensity and prevalence on the evolution of host migration. *Proceedings of the Royal Society B*, 285(1891):20182147, 2018.

[16] J Remme, O Ba, KY Dadzie, and Marc Karam. A force-of-infection model for onchocerciasis and its applications in the epidemiological evaluation of the onchocerciasis control programme in the volta river basin area. *Bulletin of the World Health Organization*, 64(5):667, 1986.

[17] Jonathan ID Hamley, Philip Milton, Martin Walker, and Maria-Gloria Basáñez. Modelling exposure heterogeneity and density dependence in onchocerciasis using a novel individual-based transmission model, epioncho-ibm: Implications for elimination and data needs. *PLoS neglected tropical diseases*, 13(12):e0007557, 2019.

[18] Maria-Gloria Basanez and Michel Boussinesq. Population biology of human onchocerciasis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 354(1384):809–826, 1999.

[19] Maria Gloria Basanez. Email, April 2023.

[20] Simon J O'Hanlon, Hannah C Slater, Robert A Cheke, Boakye A Boatin, Luc E Coffeng, Sébastien DS Pion, Michel Boussinesq, Honorat GM Zouré, Wilma A

Stolk, and María-Gloria Basáñez. Model-based geostatistical mapping of the prevalence of onchocerca volvulus in west africa. *PLoS neglected tropical diseases*, 10(1):e0004328, 2016.

[21] Elizabeth A Cromwell, Joshua CP Osborne, Thomas R Unnasch, Maria-Gloria Basáñez, Katherine M Gass, Kira A Barbre, Elex Hill, Kimberly B Johnson, Katie M Donkers, Shreya Shirude, et al. Predicting the environmental suitability for onchocerciasis in africa as an aid to elimination planning. *PLoS neglected tropical diseases*, 15(7):e0008824, 2021.

[22] AP Plaisier, GJ Van Oortmarssen, JDF Habbema, Jan Remme, and ES Alley. Onchosim: a model and computer simulation program for the transmission and control of onchocerciasis. *Computer methods and programs in biomedicine*, 31(1):43–56, 1990.

[23] World Health Organization et al. The use of onchosim for the evaluation and prediction of ocp operations-past, present and future. 1999.

[24] Luc E Coffeng, Wilma A Stolk, Honorat GM Zoure, J Lennert Veerman, Koffi B Agblewonu, Michele E Murdoch, Mounkaila Noma, Grace Fobi, Jan Hendrik Richardus, Donald AP Bundy, et al. African programme for onchocerciasis control 1995–2015: model-estimated health impact and cost. *PLoS neglected tropical diseases*, 7(1):e2032, 2013.

[25] Luc E Coffeng, Wilma A Stolk, Achim Hoerauf, Dik Habbema, Roel Bakker, Adrian D Hopkins, and Sake J de Vlas. Elimination of african onchocerciasis: modeling the impact of increasing the frequency of ivermectin mass treatment. *PloS one*, 9(12):e115886, 2014.

[26] Martin Walker, Wilma A Stolk, Matthew A Dixon, Christian Bottomley, Lamine Diawara, Mamadou O Traoré, Sake J de Vlas, and María-Gloria Basáñez. Modelling the elimination of river blindness using long-term epidemiological and programmatic data from mali and senegal. *Epidemics*, 18:4–15, 2017.

[27] Renata Retkute, Panayiota Touloupou, Maria-Gloria Basanez, T Deirdre Hollingsworth, and Simon EF Spencer. Integrating geostatistical maps and transmission models using adaptive multiple importance sampling. *medRxiv*, pages 2020–08, 2020.

[28] Olabanji Ahmed Surakat, Ayodele S Babalola, Monsuru A Adeleke, Adedapo O Adeogun, Olufunmilayo A Idowu, and Sammy O Sam-Wobo. Geospatial distribution and predictive modeling of onchocerciasis in ogun state, nigeria. *Plos one*, 18(3):e0281624, 2023.

[29] Ching-I Huang, Ronald E Crump, Emily H Crowley, Andrew Hope, Paul R Bessel, Chansy Shampa, Erick Mwamba Miaka, and Kat S Rock. A modelling assessment of short-and medium-term risks of programme interruptions for gambiense human african trypanosomiasis in the drc. *medRxiv*, pages 2022–08, 2022.

[30] Richard Selby. Video Call, April 2023.

[31] Conflict in the Democratic Republic of Congo — Global Conflict Tracker, 2023.

[32] Celia Petty. Video Call, April 2023.

[33] Ogochukwu Ibe, Obinna Onwujekwe, Benjamin Uzochukwu, Miriam Ajuba, and Paul Okonkwo. Exploring consumer perceptions and economic burden of onchocerciasis on households in enugu state, south-east nigeria. *PLoS neglected tropical diseases*, 9(11):e0004231, 2015.

[34] World Health Organization and African Programme for Onchocerciasis Control. Programme for the elimination of neglected diseases in africa (penda): strategic plan of action and indicative budget 2016-2025. Technical Report JAF19.8, 2013.

[35] Nadia al Sakkaf and Justin Alexander. Internet Access in Yemen Should Be an Opportunity for Cooperation, not a Target, 3 2022.

[36] Moses Boudourides, Andrew Stevens, Giannis Tsakonas, and Sergios Lenis. Citation graph analysis and alignment between citation adjacency and themes or topics of publications in the area of disease control through social network surveillance. In *Disease Control Through Social Network Surveillance*, pages 89–108. Springer, 2022.