Lancaster University

Masters in Data Science

School of Computing and Communications,
and Department of Mathematics and Statistics

Sub-district schistosomiasis and STH endemicity classification in Malawi

Ruotong Xiao
Student ID: 35932466

Submitted in part fulfilment of the requirements for the degree of
Masters in Data Science of the University of Lancaster,
Friday 9th September 2022

Abstract

National health departments and policy makers need more accurate disease estimates to decide whether the public health goals set have been met. Where statistical surveys are not all inclusive, spatial geographic techniques can provide more accurate disease estimates and facilitate a more comprehensive picture.

SCIF, under the guidance of WHO, has attempted to work towards controlling and eliminating the incidence of schistosomiasis and soil-transmitted helminthiasis by providing treatment and prevention of schistosomiasis and soil-transmitted helminthiasis through drugs starting in 2012. Two large-scale data collections were conducted for Malawi, the first before the provision of drugs and the second after five or six rounds of treatment. After five or six rounds of treatment, SCIF wanted to obtain a more accurate picture of disease control. Using the data collected on two occasions as a basis, this paper estimates the prevalence of schistosomiasis and soil-transmitted helminthiasis in Malawi using geostatistical techniques to obtain maps of the prevalence of each disease and to compare the effects of treatment, with confidence intervals and probabilities of exceedance.

List of figures

Figure 4.1 Impact of time on school number in Historic data.....7
Figure 4.2 Month of sample collection in Historic data 7
Figure 4.3 Shprev in different areas of historic data.....8
Figure 4.4 Smanprev in different areas of historic data.....8
Figure 4.5 anySCHprev in different areas of historic data 9
Figure 4.6 Ascprev in different areas of historic data 9
Figure 4.7 Hkwprev in different areas of historic data 9
Figure 4.8 Triprev in different areas of historic data 9
Figure 4.9 anySTHprev in different areas of historic data 10
Figure 4.10 Shprev in different areas of Reassessment data 11
Figure 4.11 Smanprev in different areas of Reassessment data 11
Figure 4.12 anySCHprev in different areas of Reassessment data 11
Figure 4.13 Ascprev in different areas of Reassessment data 11
Figure 4.14 Hkwprev in different areas of Reassessment data 12
Figure 4.15 Triprev in different areas of Reassessment data.....12
Figure 4.16 anySTHprev in different areas of Reassessment data 12
Figure 4.17 Impact of Age groups in school numbers of Historic data 13
Figure 4.18 Impact of Age groups in school numbers of Reassessment data 13
Figure 4.19 Impact of Visit date in school numbers of Reassessment data.....14
Figure 4.20 Statistics of the number of people tested for different types of diseases in Historic data 15
Figure 4.21 Statistics of the number of people tested for different types of diseases in Reassessment data 15
Figure 4.22 Statistics of the number of people positive for different types of diseases in Historic data 16
Figure 4.23 Statistics of the number of people positive for different types of diseases in Reassessment data.....16
Figure 4.24 Spatial construction for SPDEs [15].....20
Figure 4.25 Non-convex triangulated mesh to build the SPDE model.....20
Figure 4.26 Grid locations for prediction 21
Figure 4.27 Locations for prediction in Malawi 21
Figure 5.1 Map of estimated prevalence for Shprev in Reassessment data 23
Figure 5.2 Lower limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data 23
Figure 5.3 Upper limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data 24
Figure 5.4 exceedance probability of estimated prevalence for Shprev in Reassessment data.....24
Figure 5.5 Map of estimated prevalence for Smanprev in Reassessment data 25
Figure 5.6 Lower limit of 95% confidence intervals for maps of estimated

prevalence for Smanprev in Reassessment data 25
Figure 5.7 Upper limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Reassessment data 26
Figure 5.8 Exceedance probability of estimated prevalence for Smanprev in Reassessment data.....26
Figure 5.9 Map of estimated prevalence for anySCHprev in Reassessment data27
Figure 5.10 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Reassessment data.....27
Figure 5.11 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Reassessment data.....28
Figure 5.12 Exceedance probability of estimated prevalence for anySCHprev in Reassessment data.....28
Figure 5.13 Map of estimated prevalence for Ascprev in Reassessment data 29
Figure 5.14 Lower limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Reassessment data.....29
Figure 5.15 Upper limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Reassessment data.....30
Figure 5.16 Exceedance probability of estimated prevalence for Ascprev in Reassessment data.....30
Figure 5.17 Map of estimated prevalence for Hkwprev in Reassessment data....31
Figure 5.18 Lower limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Reassessment data.....31
Figure 5.19 Upper limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Reassessment data.....32
Figure 5.20 Exceedance probability of estimated prevalence for Hkwprev in Reassessment data.....32
Figure 5.21 Map of estimated prevalence for Triprev in Reassessment data.....33
Figure 5.22 Lower limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Reassessment data 33
Figure 5.23 Upper limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Reassessment data 34
Figure 5.24 Exceedance probability of estimated prevalence for Triprev in Reassessment data.....34
Figure 5.25 Map of estimated prevalence for anySTHprev in Reassessment data 35
Figure 5.26 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Reassessment data 35
Figure 5.27 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Reassessment data 36
Figure 5.28 Exceedance probability of estimated prevalence for anySTHprev in Reassessment data.....36
Figure 5.29 Map of estimated prevalence for Shprev in Historic data 37
Figure 5.30 Lower limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Historic data 37

Figure 5.31 Upper limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Historic data 37
Figure 5.32 Exceedance probability of estimated prevalence for Shprev in Historic data.....37
Figure 5.33 Map of estimated prevalence for Smanprev in Historic data 38
Figure 5.34 Lower limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Historic data 38
Figure 5.35 Upper limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Historic data 39
Figure 5.36 Exceedance probability of estimated prevalence for Smanprev in Historic data 39
Figure 5.37 Map of estimated prevalence for anySCHprev in Historic data 40
Figure 5.38 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Historic data 40
Figure 5.39 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Historic data 41
Figure 5.40 Exceedance probability of estimated prevalence for anySCHprev in Historic data 41
Figure 5.41 Map of estimated prevalence for Ascprev in Historic data 42
Figure 5.42 Lower limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Historic data 42
Figure 5.43 Upper limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Historic data 42
Figure 5.44 Exceedance probability of estimated prevalence for Ascprev in Historic data 42
Figure 5.45 Map of estimated prevalence for Hkwprev in Historic data 43
Figure 5.46 Lower limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Historic data 43
Figure 5.47 Upper limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Historic data 44
Figure 5.48 Exceedance probability of estimated prevalence for Hkwprev in Historic data 44
Figure 5.49 Map of estimated prevalence for Triprev in Historic data.....45
Figure 5.50 Lower limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Historic data.....45
Figure 5.51 Upper limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Historic data.....46
Figure 5.52 Exceedance probability of estimated prevalence for Triprev in Historic data.....45
Figure 5.53 Map of estimated prevalence for anySTHprev in Historic data 46
Figure 5.54 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Historic data.....46
Figure 5.55 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Historic data.....47

Figure 5.56 Exceedance probability of estimated prevalence for anySTHprev in Historic data 47
Figure 6.1 Map of estimated prevalence for Shprev in Reassessment data using Prevmap.....48
Figure 5.2 Standard Error for maps of estimated prevalence for Shprev in Reassessment data using Prevmap.....48
Figure 6.3 Exceedance probability of estimated prevalence for Shprev in Reassessment data using Prevmap.....49

Content

Abstract 1
List of figures 2
Content 6
1. Introduction 1
2. Background 3
3. Related work 4
4. Method 6
4.1 describing the data and data source 6
4.2 Exploratory data analyses 6
4.2.1 Missing value processing 6
4.2.2 Outlier handling 8
4.3 Data visualization 8
4.4 Explore each individual variable in the dataset 12
4.5 Model 17
4.6 SPDE 18
4.7 Mesh construction 19
4.8 Projection matrix 20
4.9 Predicted data 20
4.10 Stack with Data 22
4.11 Model formulation and model selection 22
5. Result 23
6. Discussion 48
7. Conclusion 52
Reference 53
Appendix 55

1. Introduction

Schistosomiasis is an acute and chronic parasitic disease caused by an infection that still threatens the health of hundreds of millions of people, with over 200 million people currently infected and at least 500 million at risk of infection. There are three main types of schistosomiasis, Schistosoma egypti (mainly in sub-Saharan Africa), Schistosoma japonicum (distribution limited to the People's Republic of China, Indonesia and the Philippines) and Schistosoma mansoni (mainly in sub-Saharan Africa, Brazil and the Caribbean islands). [1] Schistosomiasis has an intermediate host of freshwater snails. When such snails are infected, the larvae (caecilians) of schistosomes will be found in the water column after a period of time. When humans come into contact with infected water sources, or with infected animals, the larvae will cross the skin or water and the larvae will develop into adults in the human body, affecting the reproductive and immune systems of humans. The incubation period for schistosomiasis is about 1 to 2 months after infection and can include fever, bloody diarrhoea or blood in the urine and coughing. Schistosomiasis also has many long-term complications.

Chronic intestinal schistosomiasis manifests as debilitating symptoms such as hepatosplenomegaly (enlargement of the liver and spleen) [2,3]. Genitourinary schistosomiasis significantly affects the bladder and increases the risk of bladder cancer [4]. There is also an increased risk of HIV lesions [5]. Schistosomiasis can also lead to anaemia and malnutrition [6]. In May 2012, the World Health Assembly declared the elimination of schistosomiasis feasible in some Member States (WHA Resolution 65.21) and encouraged the use of water, sanitation and hygiene education (WASH) as part of integrated control and elimination strategies on the basis that these should reduce transmission by containing schistosome eggs and reducing human contact with water [1]. Hygiene education is important in the control of schistosomiasis because soap kills schistosome larvae and freshwater snails [6,7].

Soil-derived helminth infections (STH) are one of the most widespread infections in the world, transmitted to humans through contaminated soil and are mainly classified as roundworms, whipworms (trichuris) and hookworms (American hookworm and duodenal hookworm), of which more than 500 million people worldwide have been infected with hookworms and more than 600 million with whipworms. The main mode of transmission of soil-derived worms is through the faeces of infected people. The adult worms live in the human gut and can produce thousands of eggs a day when the gut is well nourished. These eggs can contaminate the soil and the larvae can penetrate human skin, so it is possible to be infected by walking barefoot on contaminated land. Also if manure is used as fertiliser, vegetables and fruit that have not been carefully handled and hands that have been cleaned can be infected and this disease is considered one of the most important causes of mental and physical retardation in the world. If

infected with a small number of soil-derived worms, there are usually no symptoms and only severe infections will have significant symptoms such as diarrhoea, pneumonia, intestinal bleeding, general discomfort and weakness. Soil-derived worms are usually treated with albendazole or mebendazole.

2. Background

The Malawi Ministry of Health has an ongoing national programme for schistosomiasis and soil-transmitted helminthiasis throughout the country. The plan is to start treating schistosomiasis with praziquantel (PZQ) and soil-transmitted helminthiasis with albendazole (ALB) on an annual basis nationwide in 2012. According to WHO guidelines, after five to six rounds of treatment, an evaluation should be conducted to reassess the frequency of treatment and to ensure that the programme focuses resources where the need is greatest. SCIF therefore supported baseline mapping in 2012-13, with all 29 districts being endemic areas for Schistosoma mansoni and Schistosoma haematobium. SCIF supported a national reassessment survey during 2017-2019. In both the initial mapping and reassessment, the assessment area was district.

The drugs needed for the national treatment campaign were donated through the WHO drug donation programme. In support of the new WHO guidelines for the control and elimination of schistosomiasis, the WHO Regional Office for Africa requires countries in Africa to submit drug requests based on sub-regional administrative units. Few countries have sufficient infection data at this sub-district level to make accurate prevalence estimates needed for programme decisions.

The project will support the Malawi Ministry of Health (MoH) to update its national strategy for schistosomiasis and soil-transmitted helminthiasis (STH) in line with the WHO guidelines and to achieve sub-district implementation. The sub-district level in Malawi is the traditional authority level.

3. Related work

About Schistosomiasis in Malawi, published by Peter Macolla in 2014: The Systematic Review provides a good summary of the work from 1985 to 2014. The systematic review discusses the environment of schistosomiasis in Malawi. The Systematic review argues that unsafe water and lack of sanitation play a major role in the transmission of schistosomiasis in Malawi. Lake Malawi, also a major tourist destination, is often cited as a possible source of schistosomiasis infection.[1]

Ying-SiLai and Patricia Biedermann in 2015 conducted a meta-analysis for high-resolution risk assessment of Schistosoma spp. infections and praziquantel treatment doses by Bayesian ground truth logistic regression models for 44 countries/regions in sub-Saharan Africa at different administrative levels, predicting the risk of infection on a 1,155,818-pixel grid at 5 × 5 km based on environmental and socio-economic predictors and calculating the praziquantel dose needed to prevent morbidity [18].

MH Hodges, RJ Soares Magalhães, J Paye conducted the first comprehensive national mapping of urogenital schistosomiasis in Sierra Leone in 2012, a study that made joint spatial predictions of schistosomiasis and soil-derived helminths, again in 2012 to estimate disease infection rates prior to the use of drug treatment, set of predictor variables included population density, LST, NDVI, PIWB and elevation. to provide the first map of genitourinary schistosomiasis distribution in the country. In addition, by combining this map with previously reported maps of intestinal schistosomiasis and STH, the population at risk of school-age children requiring an integrated schistosomiasis/STH treatment programme was estimated based on the homogeneity of these diseases [19].

UF Ekpo and E Hürlimann in 2013 studied the prevalence of schistosomiasis in Nigeria by fitting a Bayesian geostatistical model with a Markov chain Monte Carlo (MCMC) simulation method to S. haematobium infection risk at unobserved sites by combined Bayesian kriging. Point prevalence maps and the amount of praziquantel required for treatment were obtained [20].

In 2011, Nadine Schur and Eveline Hürlimann used a Bayesian geostatistical model based on environmental predictors and a joint kriging method to create smoothed empirical prevalence maps for Schistosoma mansoni and Schistosoma egypti in people ≤20 years of age in West Africa (including Cameroon), as well as to derive country-specific prevalence estimates [21].

In 2009, G. RASO and B. MATTHYS examined risk factors for Schistosoma mansoni infection among school-age children in the Manzan Mountains of western Côte d'Ivoire,

using logistic ground statistical models and Bayesian kriging for spatial risk prediction and mapping at non-sampled sites [22].

4. Method

4.1 describing the data and data source

The data sources are twofold, with SCIF providing data from both surveys and the Malawian traditional authority level administrative classification. The data from both surveys are in csv format and contain the year of the survey, the month of the survey, the district of the survey site, the name of the school, the longitude and latitude of the school, the type of reagents tested, the maximum and minimum age of the students tested, the two species of schistosomiasis and the three types of soil-transmitted helminthiasis diseases. The traditional authority level for administrative classification in Malawi is the sub-district level, and the two tests essentially ensure that the district level is tested, but there are many areas under the sub-district level that are not detected, and SCIF hopes that geostatistical techniques will allow accurate prevalence estimates to be made.
Altitude data were obtained from the digital elevation model (DEM)from the Shuttle Radar Topography Mission(SRTM).

4.2 Exploratory data analyses

4.2.1 Missing value processing

Missing values in the data were first dealt with; of the two measurements, the data from 2013-2015 (the first survey) had a relatively large number of missing values, mainly in the temporal data from the first survey, where most tests did not record the time at that time, so the most recent available data was used to fill in.

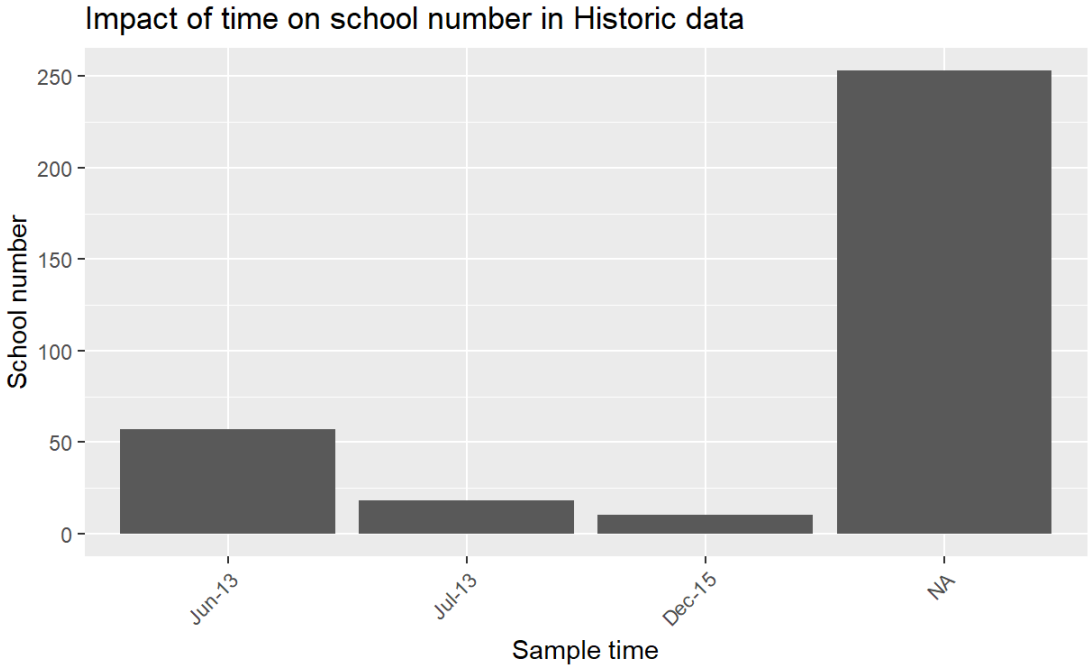


Figure 4.1 Impact of time on school number in Historic data

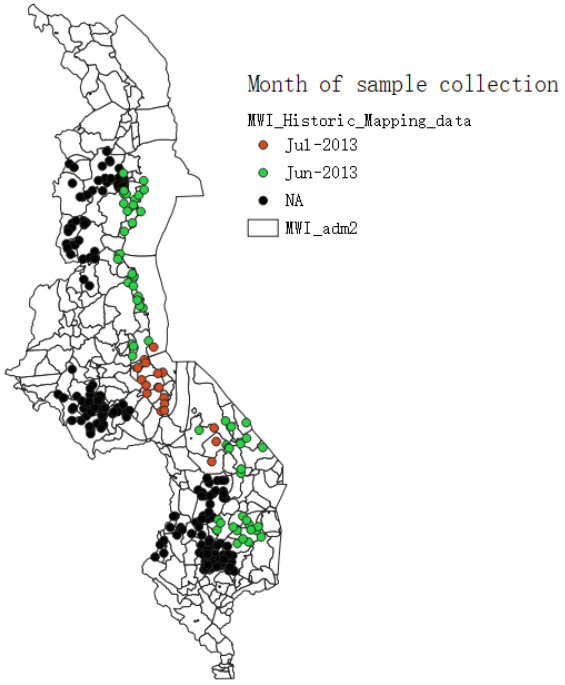


Figure 4.2 Month of sample collection in Historic data

4.2.2 Outlier handling

Most of the outliers were latitude and longitude that were not within the Malawi country, some of these were filled in backwards and others were adjusted to still be outside the map and this data was removed. The data detected in December 15 all occurred on a small island in Lake Malawi, which is very small and not part of the mission, so this data was removed. There were also significant outliers in the data for the number of people detected, but these were reasonable and were not changed.

4.3 Data visualization

In each of the two surveys, two SCH and three STH diseases were investigated, along with the sum of the infection rates for the two SCH diseases and the sum of the infection rates for the three STH diseases. The main task of this paper is to generate statistical plots of the prevalence of each disease at the traditional authority level, so first we show the infection rates for the raw data. The following seven graphs are the results from the first survey.

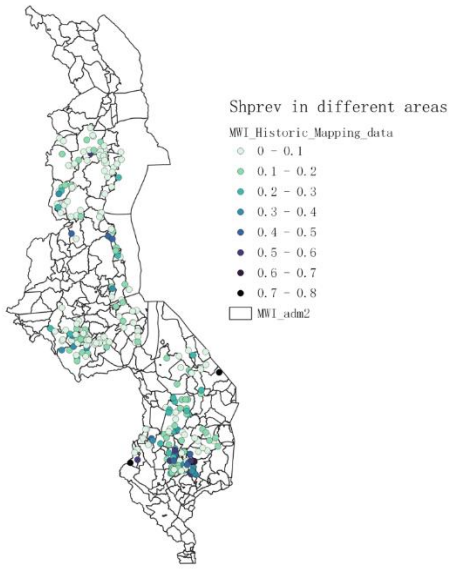


Figure 4.3 Shprev in different areas of historic data

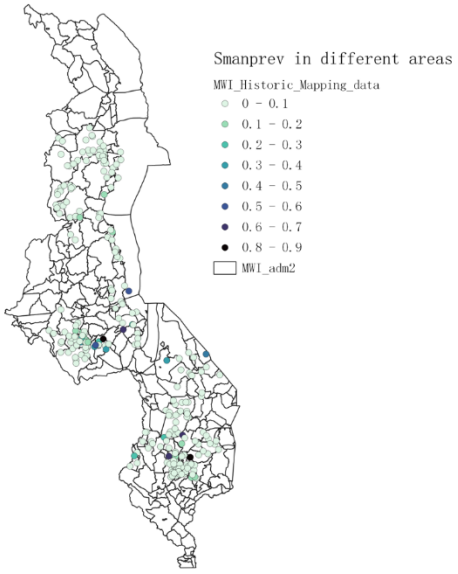


Figure 4.4 Smanprev in different areas of historic data

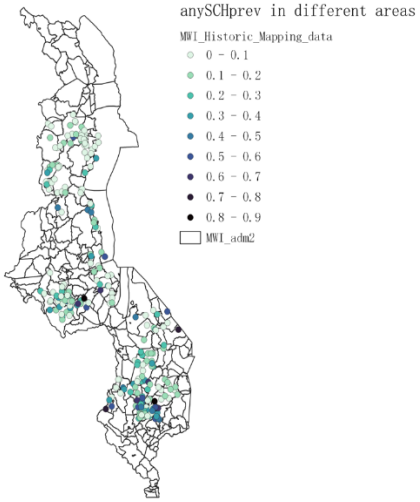


Figure 4.5 anySCHprev in different areas of historic data

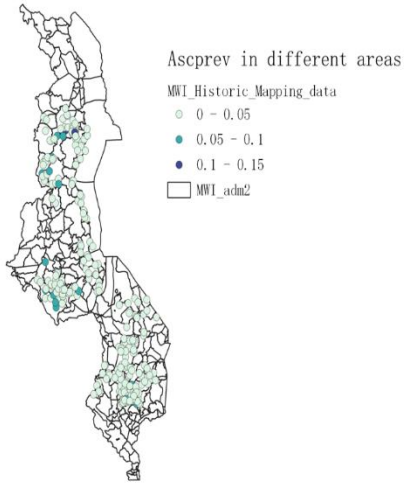


Figure 4.6 Ascprev in different areas of historic data

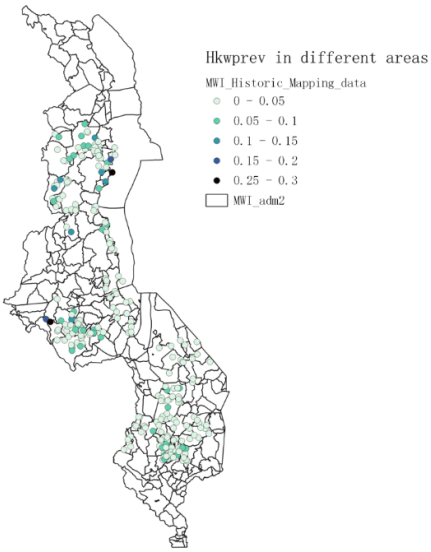


Figure 4.7 Hkwprev in different areas of historic data

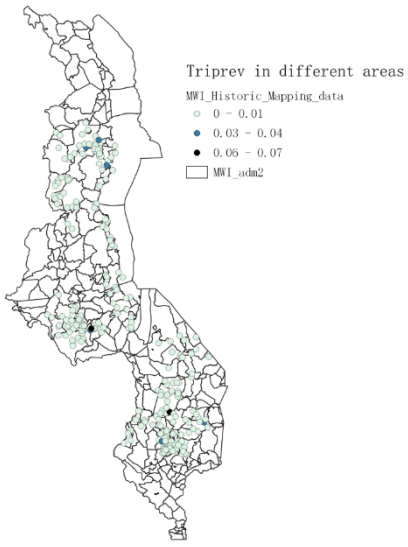


Figure 4.8 Triprev in different areas of historic data

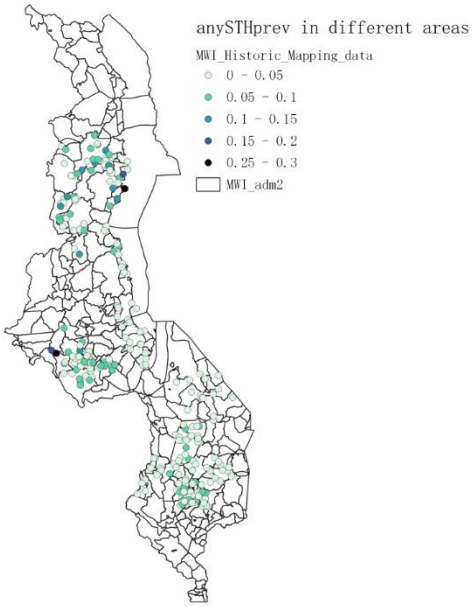


Figure 4.9 anySTHprev in different areas of historic data

The results of this second survey show that the data are more numerous and more evenly distributed, and that there has been a significant decrease in the rate of infection in Sman, suggesting that the treatment has had an effect over time. At the same time, the rate of STH infection has increased rather than decreased, especially in Asc, where the disease is still spreading despite the intervention.

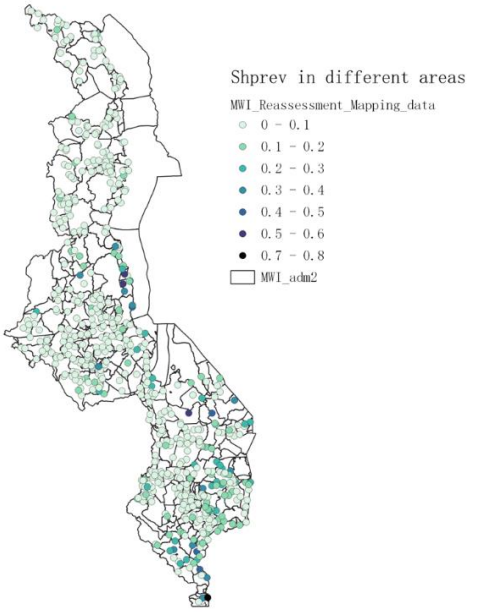


Figure 4.10 Shprev in different areas of Reassessment data

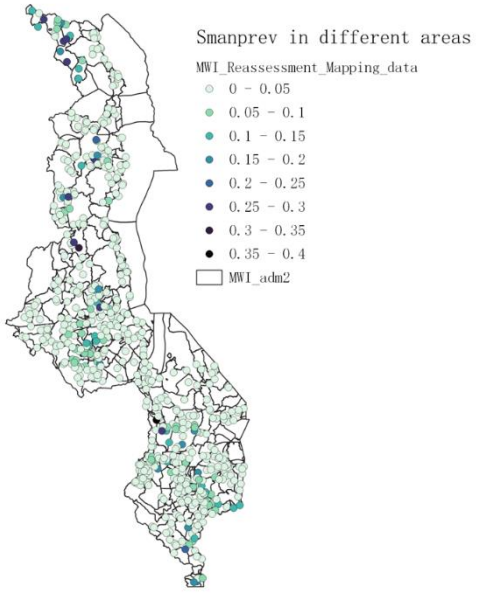


Figure 4.11 Smanprev in different areas of Reassessment data

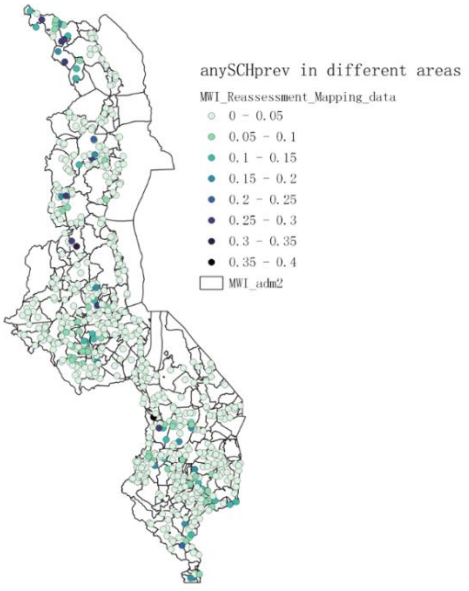


Figure 4.12 anySCHprev in different areas of Reassessment data

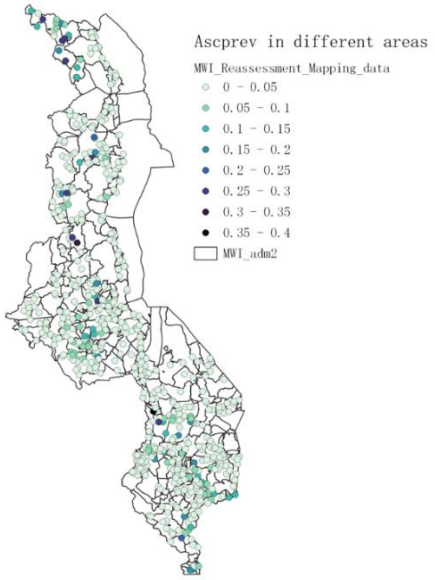


Figure 4.13 Ascprev in different areas of Reassessment data

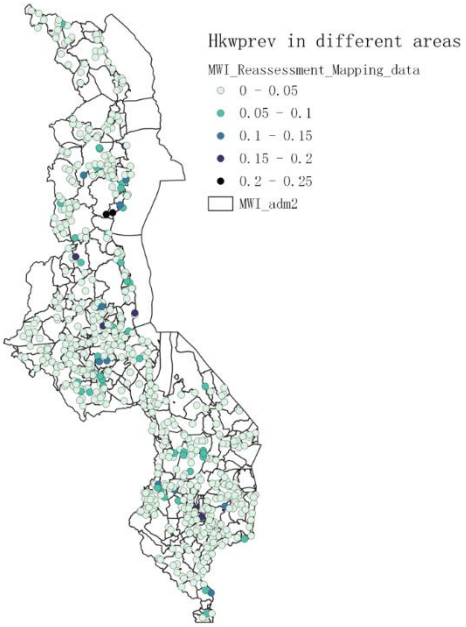


Figure 4.14 Hkwprev in different areas of Reassessment data

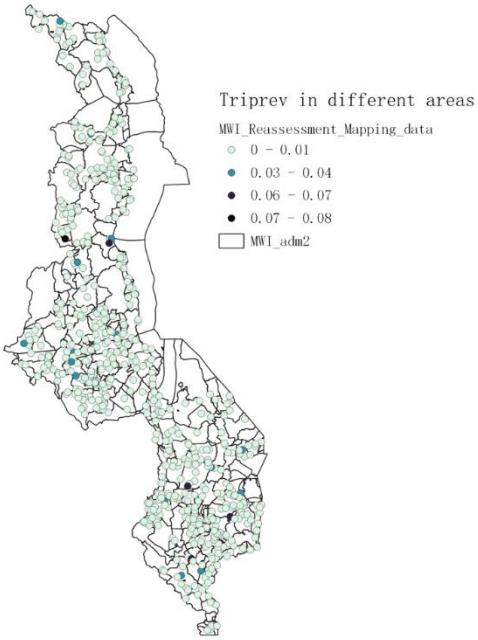


Figure 4.15 Triprev in different areas of Reassessment data

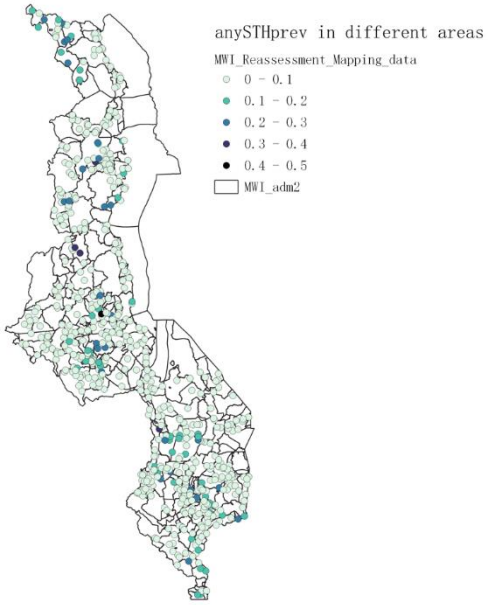


Figure 4.16 anySTHprev in different areas of Reassessment data

4.4 Explore each individual variable in the dataset

Due to the small number of data variables and the small total amount of data, each variable was analysed step by step. The age range of those surveyed at the time of both surveys is shown here and it can be seen that in both surveys 10 to 14 years was clearly the dominant age, followed by 10 to 13 years. Also the average age of the first survey is older and the five-year gap between the two surveys allows for the conclusion that there is essentially no replication of the experiment

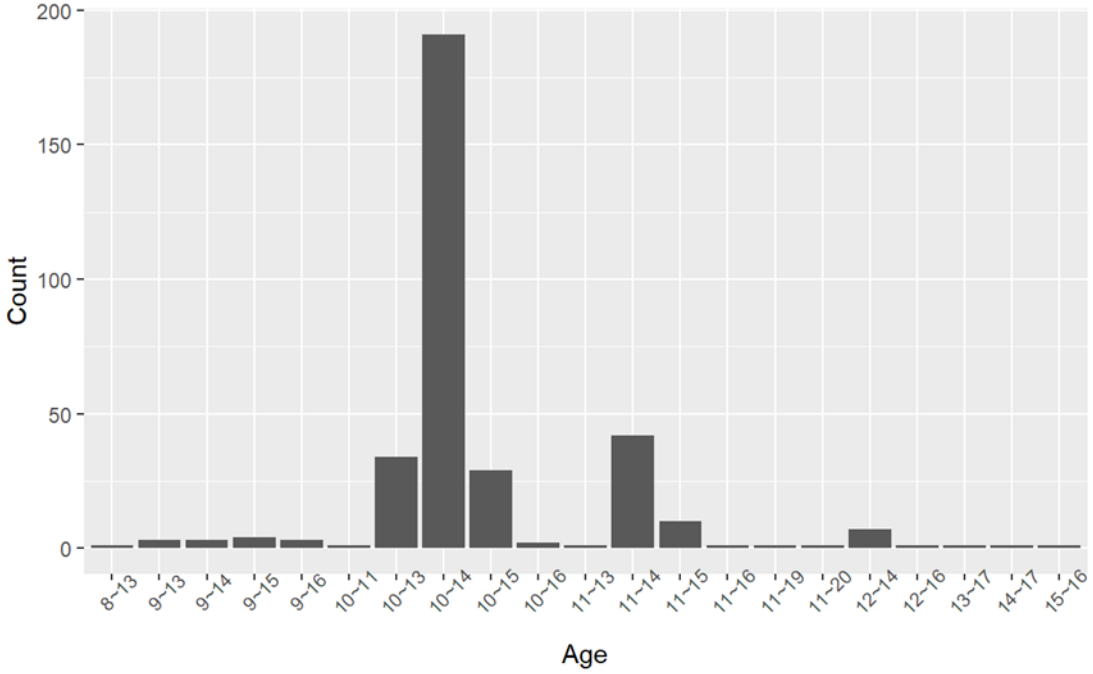


Figure 4.17 Impact of Age groups in school numbers of Historic data

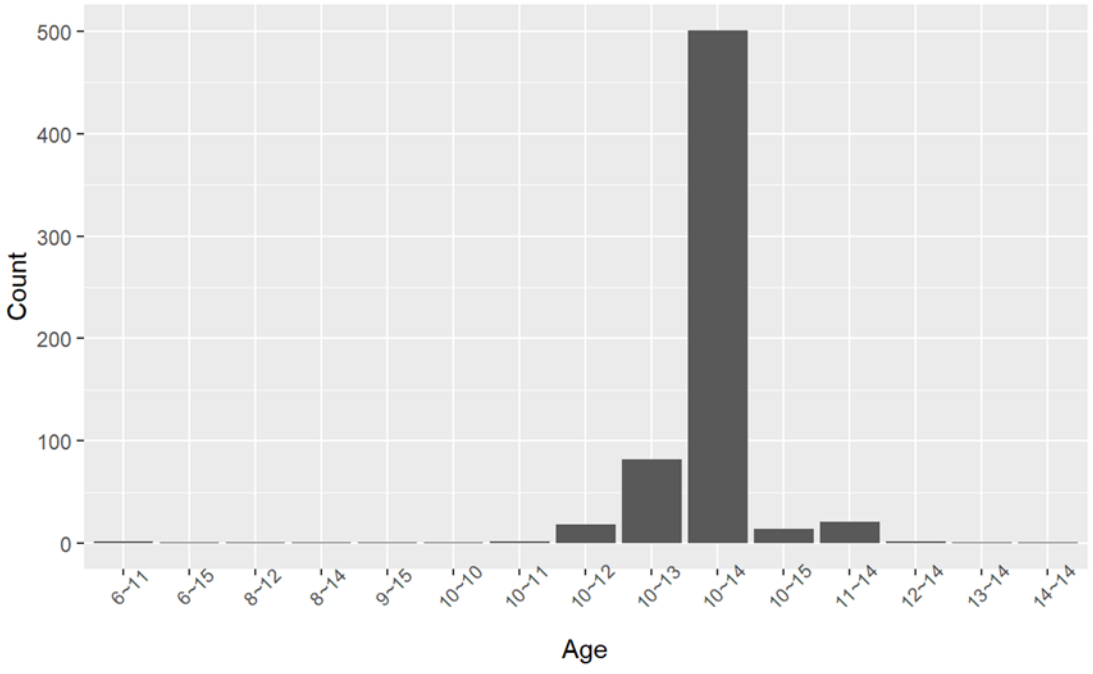


Figure 4.18 Impact of Age groups in school numbers of Reassessment data

The statistics for the month of the second survey are shown here, which are of less research value due to the large number and relative concentration of missing values in the first survey. It can be seen that the statistics are concentrated in spring and summer.

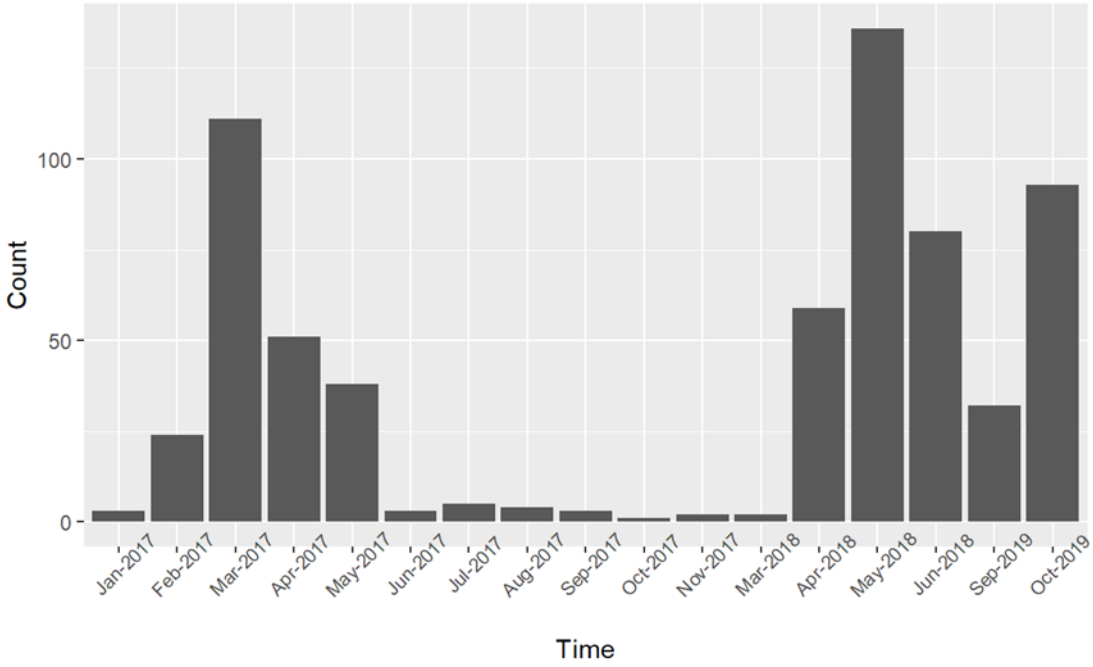


Figure 4.19 Impact of Visit date in school numbers of Reassessment data

The two box plots show the number of people tested for each disease in both surveys, and it can be seen that essentially all are clustered around 30 people, the box plots are compressed into a single line with fewer outliers, and also by retrieving the data it was obtained that most of the outliers came from the same tests, i.e. the vast majority of the tests tested the same number of people on all five diseases.

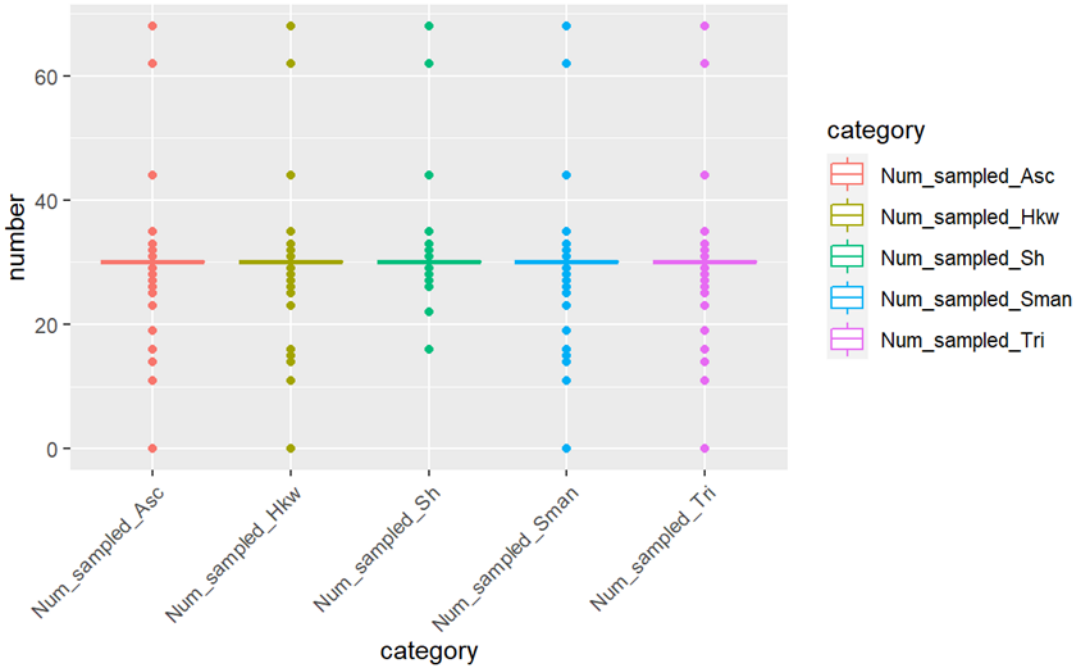


Figure 4.20 Statistics of the number of people tested for different types of diseases in Historic data

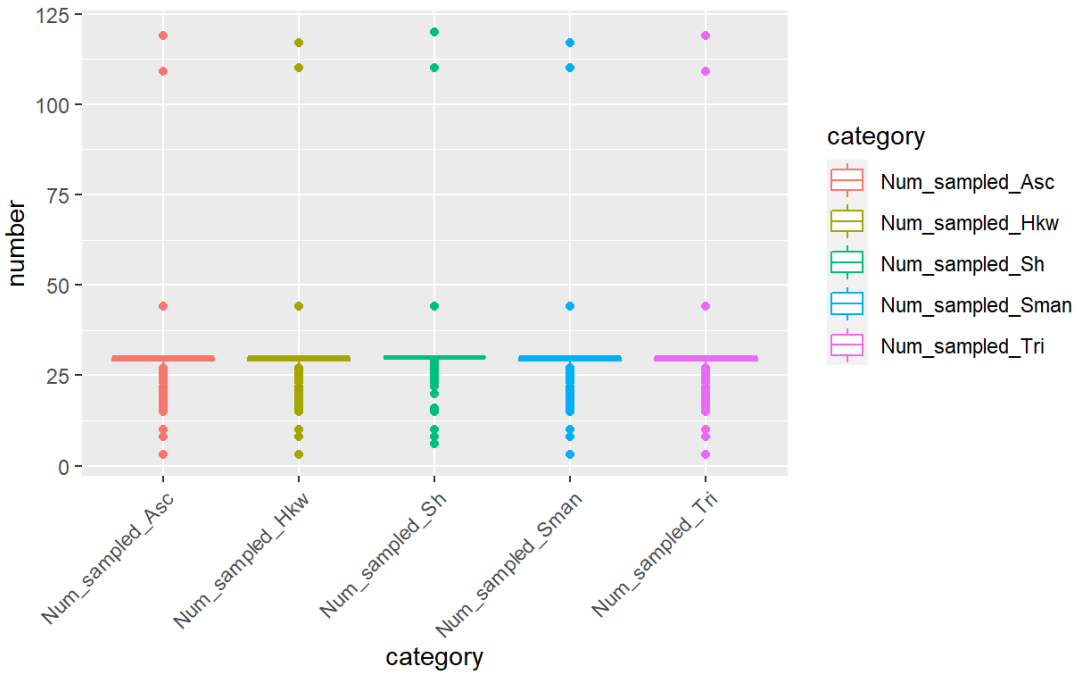


Figure 4.21 Statistics of the number of people tested for different types of diseases in Reassessment data

These two graphs show the distribution of the number of infections for the five diseases between the two surveys. It can be seen that, with the exception of Asc, the other four

diseases show a decreasing trend in outliers, where basically one point represents a school, and as the values decrease in both surveys they are also accompanied by an increase in outliers.

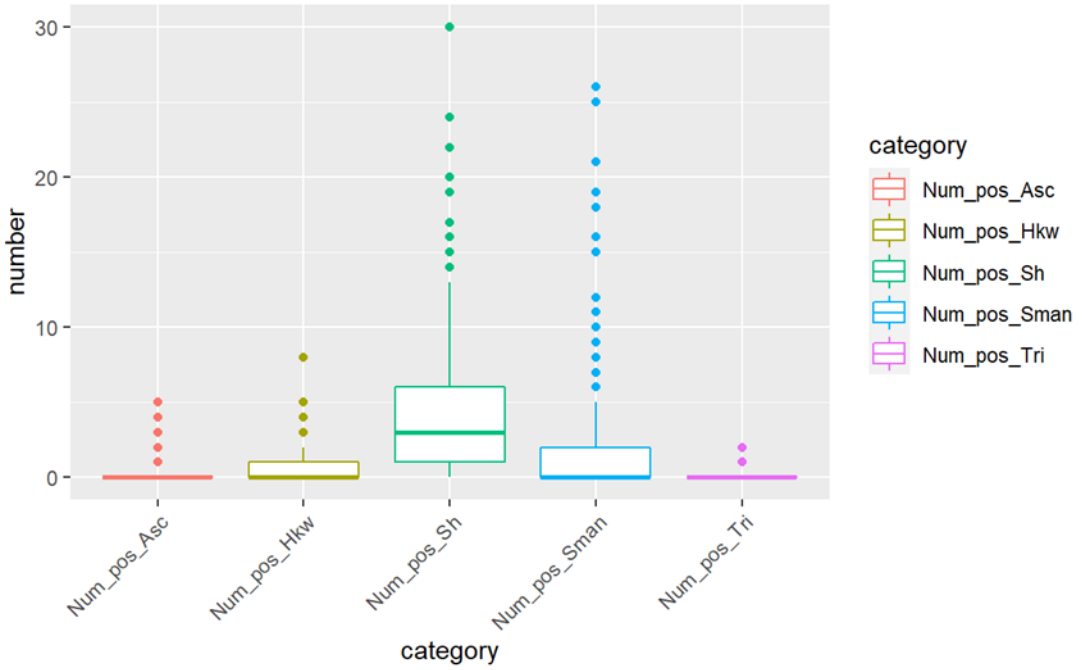


Figure 4.22 Statistics of the number of people positive for different types of diseases in Historic data

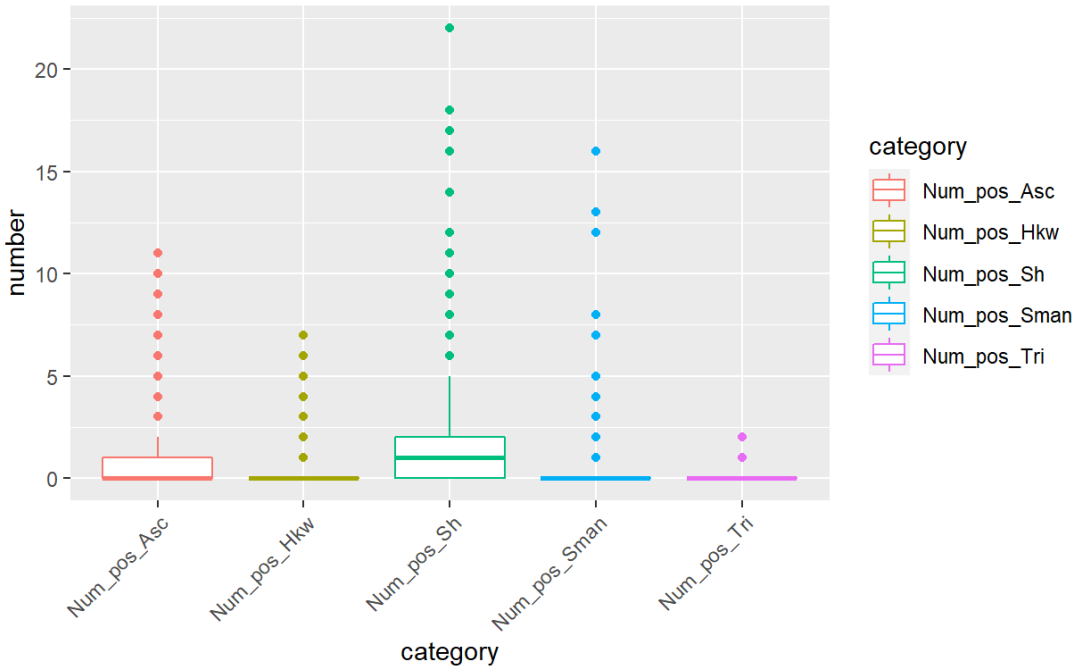


Figure 4.23 Statistics of the number of people positive for different types of diseases in Reassessment data

4.5 Model

Disease mapping is a fundamental task in spatial epidemiology, and its main purpose is to present spatial or temporal variation in disease risk on a map [9], providing researchers with a more intuitive perception and facilitating other studies. Traditional disease maps are based on administrative boundaries plotting estimated crude rates discrete in space, such as punctuated maps and equivalence area maps [10], an approach that is prone to problems such as visual bias due to differences in the shape and size of individual areas. Researchers now generally agree that a more appropriate approach is to map disease using spatial statistics, with the main methods being geostatistics and Bayesian statistics. The basic idea behind these methods is to use methods such as 'interpolation' or 'smoothing' to process coarse rate estimates to produce spatially continuous and smooth maps of disease that are easy to interpret [10].

Spatial autocorrelation is a common problem involved in geographical research, as Toblers law states, Things that close together and likelier to be the same, whether in time or space, are present, and every disease is affected by both time and space. As we know, in epidemiology, spatial analysis is mainly about describing disease data spatially and temporally, finding correlations and mapping disease risk, but in practice spatial analysis is very complex, computationally intensive and not easy to visualise [11]. INLA package gives a simple, easy and efficient solution to spatial analysis.

For the disease mapping problem,

Consider n independent replications of an experiment at location $s_i, i \in [1, n]$, in this study, the replications are tests of individuals for a specific disease. The results of the experiment *Yi* means the number of students test positive for schistosomiasis or soil-transmitted helminthiasis, the value should satisfy

$$Y_i \sim Binomial\left(p(s_i), N_i\right) \tag{3.1}$$

where *Ni* is the number of experiments, while $p(s_i)$ is the probability of positive for location s_i, This formula indicates that the model is considered to be a binomial distribution model, which is a discrete probability distribution of the number of successes in *n* independent successful/failed trials,

Here a logarithmic equation is used to relate the probabilities in the range [0, 1] to the existing variables, $p(s_i)$ satisfies(si)

$$\text{log}it\left(p(s_i)\right) = \beta_0 + X(s_i)\beta + \psi(s_i) \tag{3.2}$$

where $\beta_0$is the intercept and β is the covariance coefficient (parameter) and $\psi(s_i)$ is

a Gaussian field that can be approximated using a Gaussian Markov Random Field (GMRF), which can be shown as

$$\psi(s_i) \sim N(0, \Sigma) \tag{3.3}$$

Σ is the covariance matrix, determined by the spatial correlation function. In this paper we will use the Matern function [12].

$$Cov(Z(s_i), Z(s_j)) = \frac{\sigma^2}{\Gamma(\lambda)\, 2^{\lambda-1}} (\kappa \| s_i - s_j \|)^{\lambda} K_{\lambda}(\kappa \| s_i - s_j \|) \tag{3.4}$$

si sj is two different location,$\Gamma(\ )$ is Gamma function,$K_{\lambda}(\ )$ is the Bessel function of the second kind, κ, λ are nonnegative parameters of covariance,λ is usually equal to 0.5. It is often used to define the covariance between two measurements. Since the covariance depends only on the distance between two points, it is stationary.

Gaussian Markov random fields

Markov random fields provide a model generally used to express the interaction between spatially relevant random variables by introducing structural information through a properly defined system of neighbourhoods, and are based on MRF models and Bayes estimation to determine the solution to a problem according to optimal criteria in statistical decision and estimation theory [14]. Gaussian Markov random fields (GMRFs) are Markov random fields that satisfy a Gaussian distribution, are probabilistic graphical models widely used in spatial statistics and related fields to model dependencies over spatial structures [13]. Compared with another type of Gibbs distribution Markov random field, Gaussian Markov random fields are less computationally intensive and have wider applications.

4.6 SPDE

Stochastic partial differential equations (SPDEs) are essentially partial differential equations with random terms and random coefficients, and when considering geostatistical data, we can usually assume the existence of a spatially continuous variable for the observations that can be modelled using a Gaussian random field [15]. We can then use the stochastic partial differential equation approach implemented in the R-INLA package to fit the spatial model to predict variables at unsampled locations (Lindgren and Rue 2015). The Gaussian random field using the Matérn covariance matrix can be expressed as a solution to the following continuous domain stochastic partial differential equations [15].

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} (\tau x(s)) = W(s) \tag{3.5}$$

In this equation,κ isa scale parameter,Δ is the Laplacian defined by

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} \tag{3.6}$$

d is the dimension of the spatial domain, α is the parameter that controls the smoothness of the Gaussian random field, τ is the parameter that controls the variance, $x(s)$ is a Gaussian random field and $W(s)$ is a Gaussian spatial white noise process [15].

Parameters of the Mattern covariance function and SPDE is link by $\lambda = \alpha - \frac{d}{2}$

$$\sigma^2 = \frac{\Gamma(\lambda)}{(4\pi)^{d/2}\Gamma(\alpha)\kappa^{2\lambda}\tau^2} \tag{3.7}$$

An approximate solution of the SPDE can be found using the finite element method by dividing the spatial domain D into a network of disjoint triangles, a mesh having n nodes and n basis functions.

The basis functions $\psi_k(x)$ is defined as a segmented linear function on each triangle, equal to 1 at the vertices k and equal to 0 at the other vertices. then, the continuously indexed Gaussian field x is then expressed as a discrete indexed Gaussian Markov random field (GMRF) via a finite basis function defined on the triangularised mesh [15].

$$x(s) = \sum_{k=1}^{n} \psi_k(s) x_k \tag{3.8}$$

In this eqution,n is the number of node(vertices of the triangulation), $\psi_k(x)$ denotes the segmented linear basis functions, and x_k represents the weights of the Gaussian distribution. This Gaussian distribution has a mean of zero.

4.7 Mesh construction

The SPDE method approximates the continuous Gaussian field Z(·) as a discrete Gaussian Markov random field through a finite basis function defined on a triangularised grid of the study area. To avoid boundary effects, which is a phenomenon where the outermost data is subject to large errors due to the absence of data in the vicinity, the grid needs to cover the study region as well as the outer extensions, thus increasing the variance near the boundary. [12] There are no defined rules governing the size of the grid, a large number of grids consumes more computational power and

the model is not necessarily better than a larger grid. Therefore, in order to save computational resources, small triangles are generally used in the region and large triangles outside the region. Also, to prevent the construction of many small triangles in a very close region, a minimum allowable distance is set. Figure 4.25 is a example based on the data of Malawi.

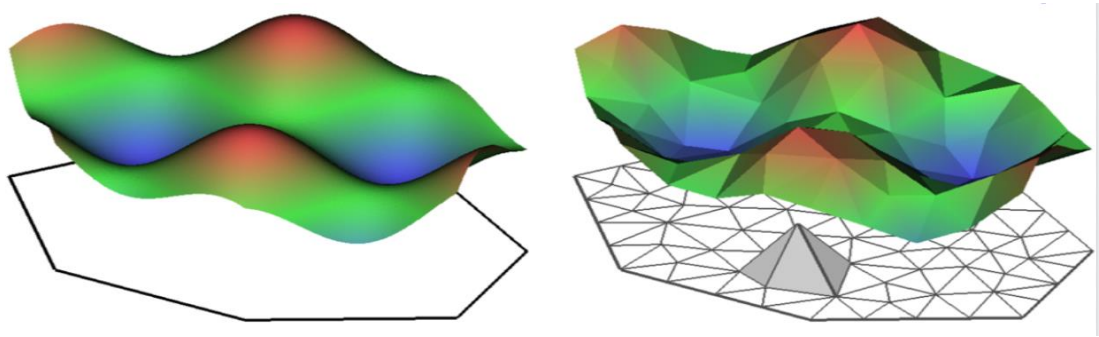


Figure 4.24 Spatial construction for SPDEs [15]

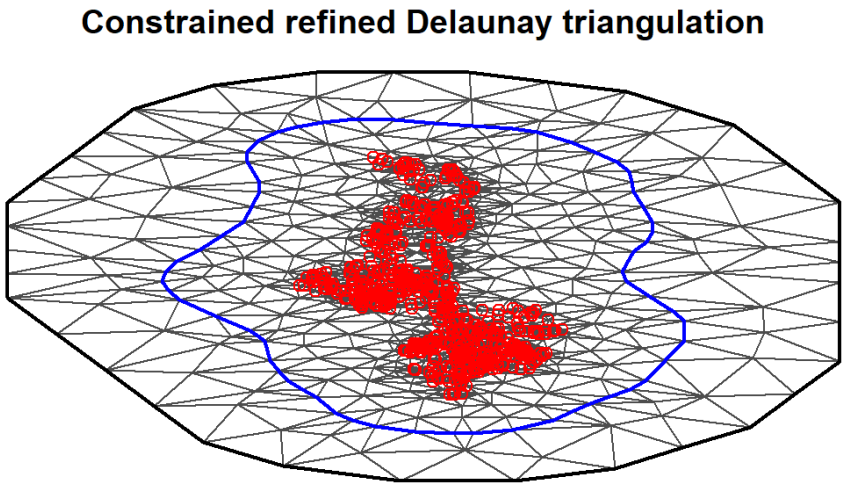


Figure 4.25 Non-convex triangulated mesh to build the SPDE model

4.8 Projection matrix

The next stage is constructing a projection matrix to project the spatial-temporal continuous Gaussian random field from the observations to the grid nodes. This step allows the existing observations to be combined with the already constructed network structure, so that the projection matrix has a number of rows equal to the number of observations and a number of columns equal to the number of triangular vertices of the mesh structure.

4.9 Predicted data

The next is construct a matrix that covers exactly all the predicted positions, requiring

each circle to be close to each other but not covered at the same time, this is to prevent multiple predictions being made at the same position. Here is a sample using the data of Malawi.

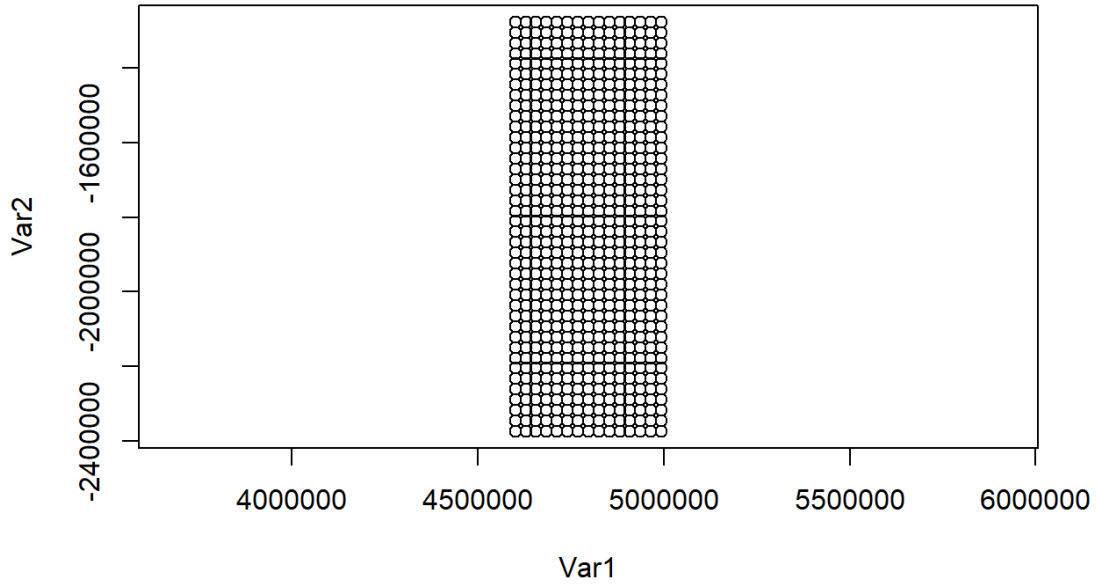


Figure 4.26 Grid locations for prediction

Only the required portion is then retained, and this shape should exactly match the contours of the Malawi map. Each circle in this matrix will yield a value for the predicted outcome, and stitching these together creates the final statistical map of infection rates.

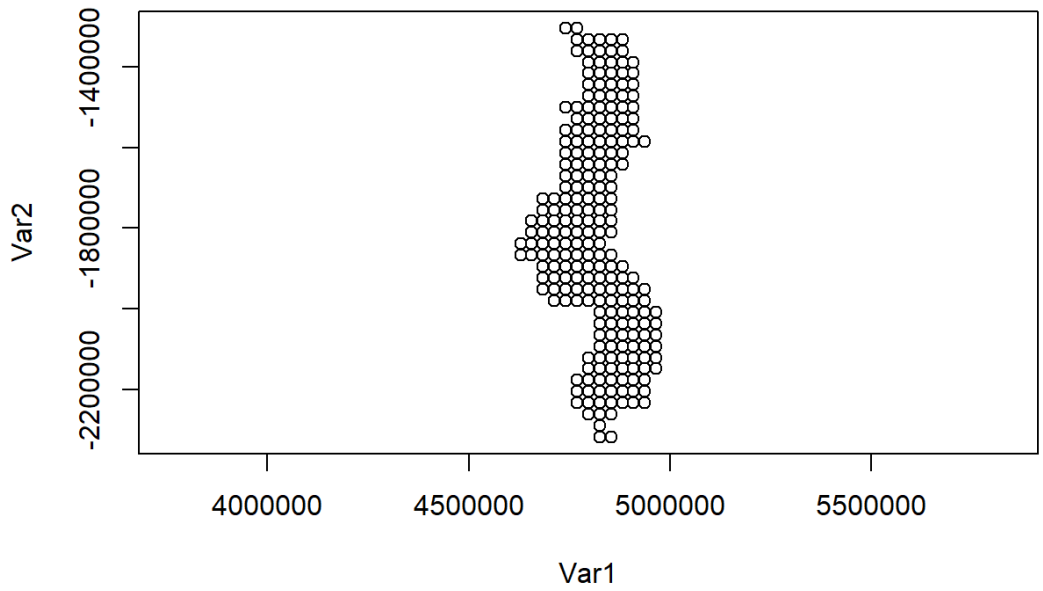


Figure 4.27 Locations for prediction in Malawi

4.10 Stack with Data

Since the covariates have already been evaluated at the observation location, we only want to apply the projection matrix to the spatial effects, not the fixed effects. Performing this manually is difficult, but the inla.stack. function does the trick. Think of it as creating a list of the terms needed to build the model. The three main parameters are a list of vectors containing data (data), a list of projector matrices (each matrix is associated with a block effect A) and a list of effects (effects). (Optionally) labels can be assigned to the data stack (using parameter labels) [15].

4.11 Model formulation and model selection

The altitude variable and the forest cover variable were downloaded from the open-source database at the time of modelling, in addition to the variables already present in the dataset, due to the large time span of sampling at different times. It was difficult to obtain accurate forest cover at the time of measurement, so this variable was eventually dropped. Other variables were year, elevation, maximum student age, and minimum student age. The lack of analysis of month data is an area where this paper could be improved, as the model always has problems with the inclusion of time series due to the limited ability of the code.

In INLA, there are metrics to assist in model selection, including DIC, CPO/PIT and WAIC, and the model with the lowest WAIC is chosen in this paper.
The model ultimately chose the maximum student age, minimum student age and altitude variables as the WAIC values were the smallest at this point

5. Result

The results are presented in four graphs, the first showing the estimated prevalence of shprev at the second survey, the second showing the lower limit of the 95% confidence interval, the third showing the upper limit of the 95% confidence interval, and the fourth showing the 20% excess probability, i.e. the probability that the infection rate exceeds 20%. It can be seen that the southern and central parts of the shprev are still at higher infection rates after treatment, with the most southern areas having the highest infection rates, and this model was selected several times and has the smallest 95% confidence interval range of all the models attempted. The overall change from the first survey was significant, with significant decreases in infection rates around Mzuzu, Blantyre and Lilongwe.

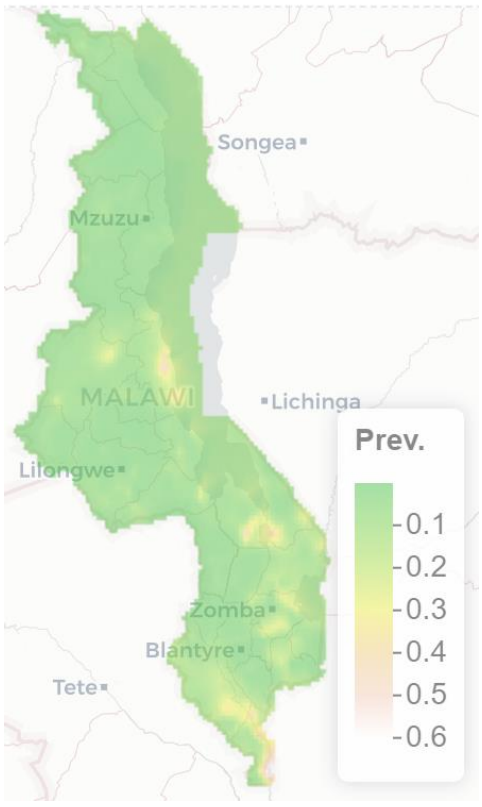


Figure 5.1 Map of estimated prevalence for Shprev in Reassessment data

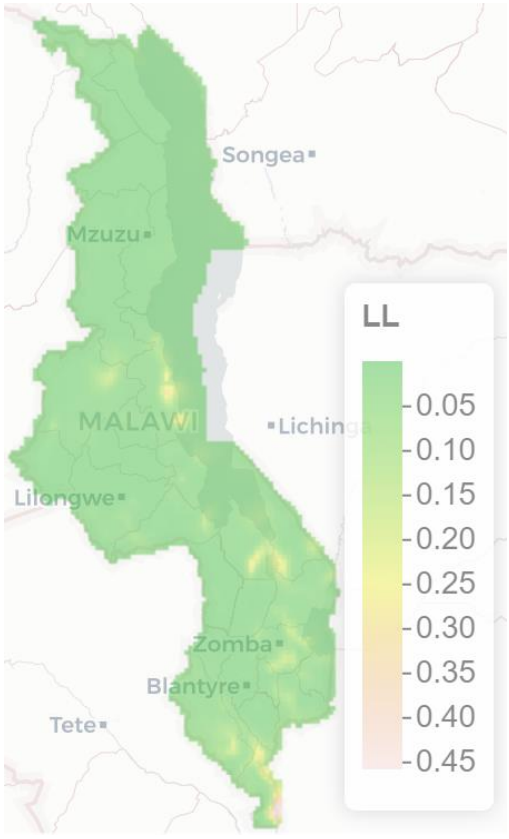


Figure 5.2 Lower limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data

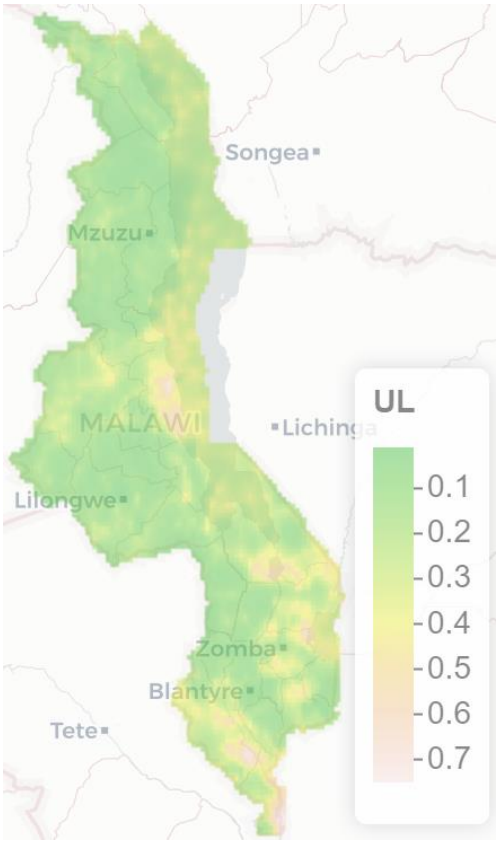


Figure 5.3 Upper limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data

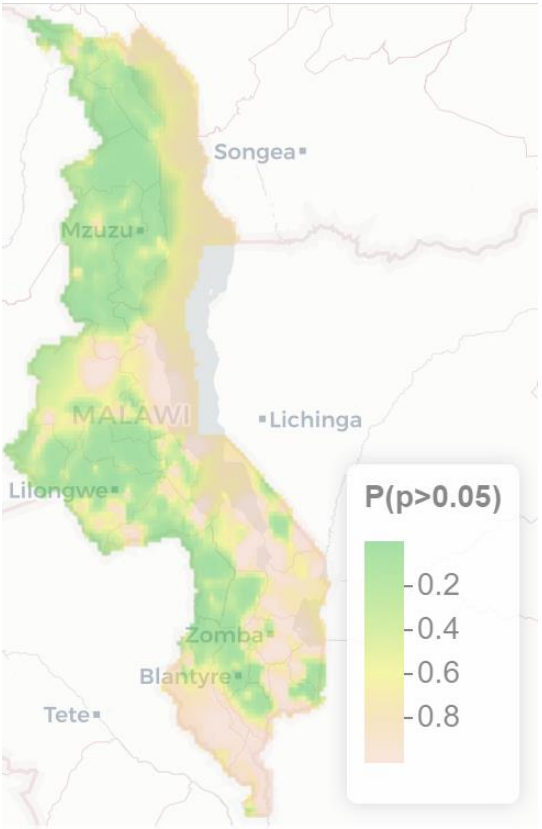


Figure 5.4 exceedance probability of estimated prevalence for Shprev in Reassessment data

The four graphs show the results of Sman's second survey. It can be seen that, as with Sh, there are still some high rates of prevalence in the south and central regions after treatment, with the highest rates in the central region. The overall change from the first survey is not very significant, but the overall rate of prevalence has decreased.

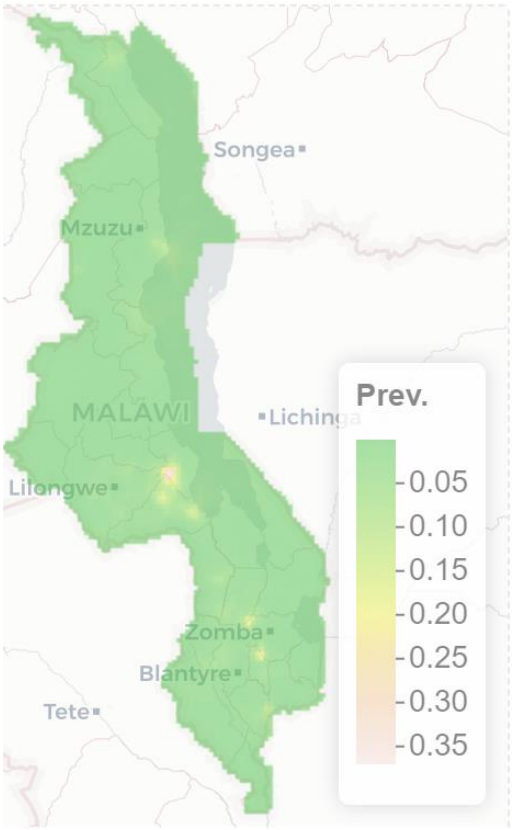


Figure 5.5 Map of estimated prevalence for Smanprev in Reassessment data

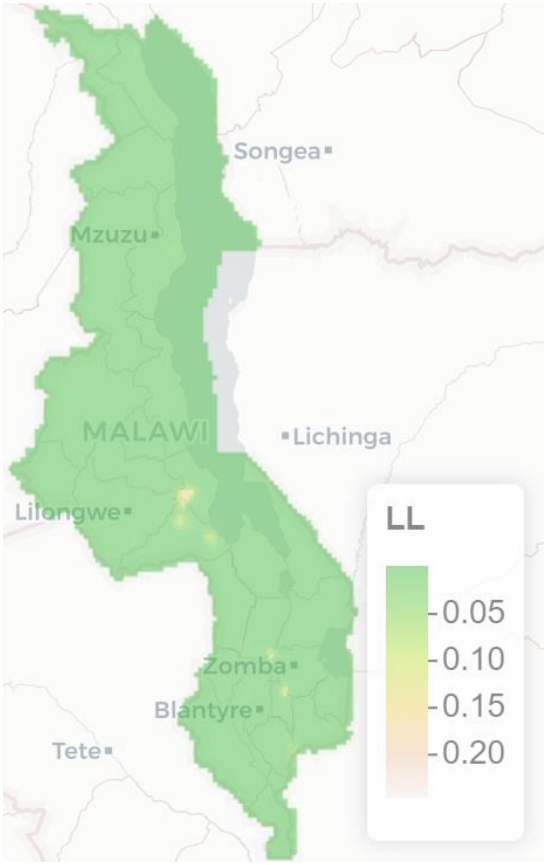


Figure 5.6 Lower limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Reassessment data

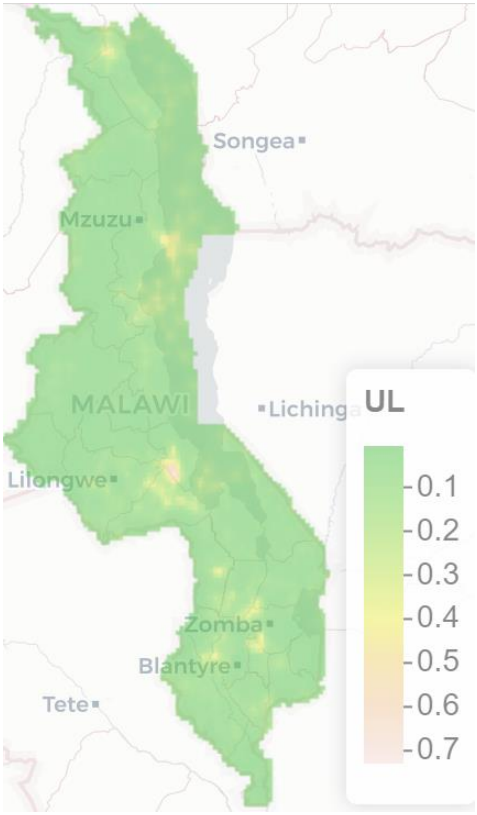


Figure 5.7 Upper limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Reassessment data

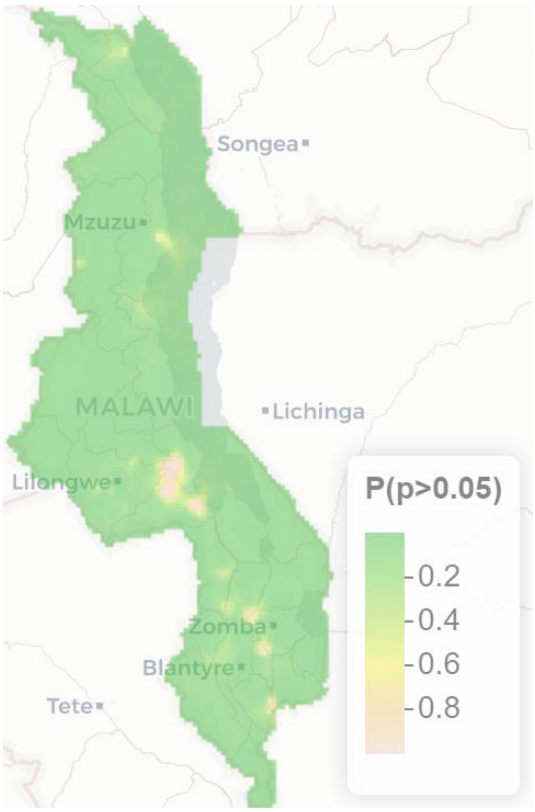


Figure 5.8 Exceedance probability of estimated prevalence for Smanprev in Reassessment data

The following four graphs show the results for Sman and sh and at the second survey. It can be seen that some of the southern and central areas are still at high prevalence after treatment, with the highest prevalence in the southernmost areas. The overall change in the results from the first survey is clear, with the overall prevalence of infection slowing down

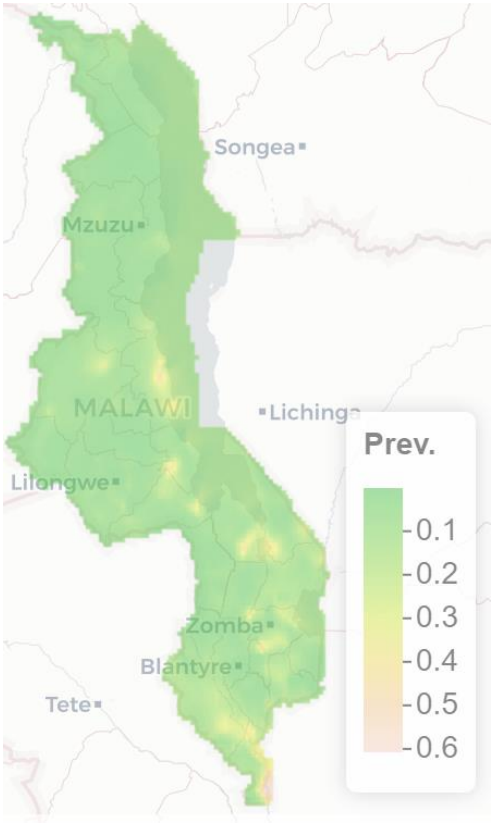


Figure 5.9 Map of estimated prevalence for anySCHprev in Reassessment data

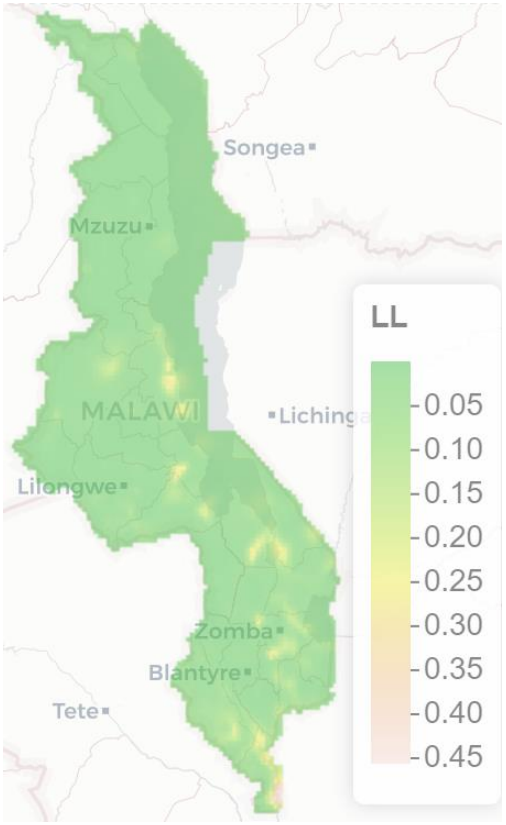


Figure 5.10 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Reassessment data

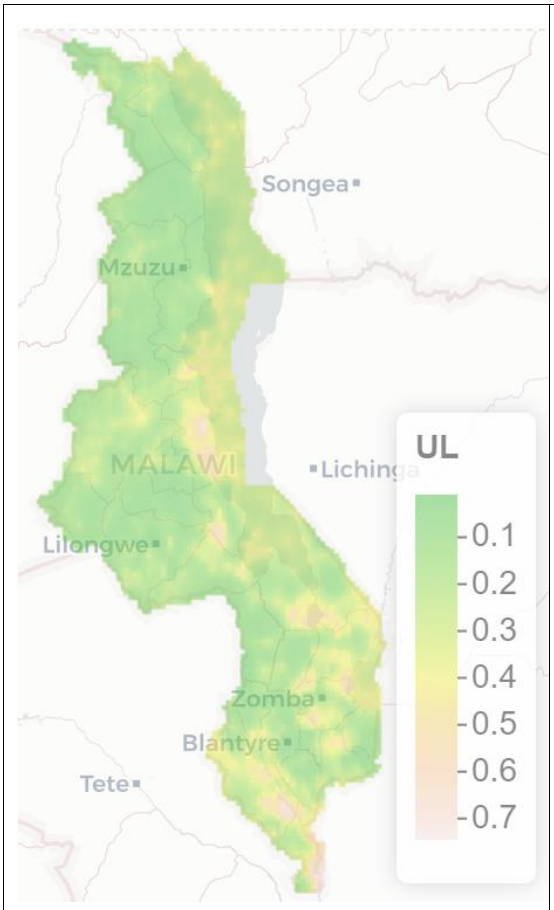


Figure 5.11 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Reassessment data

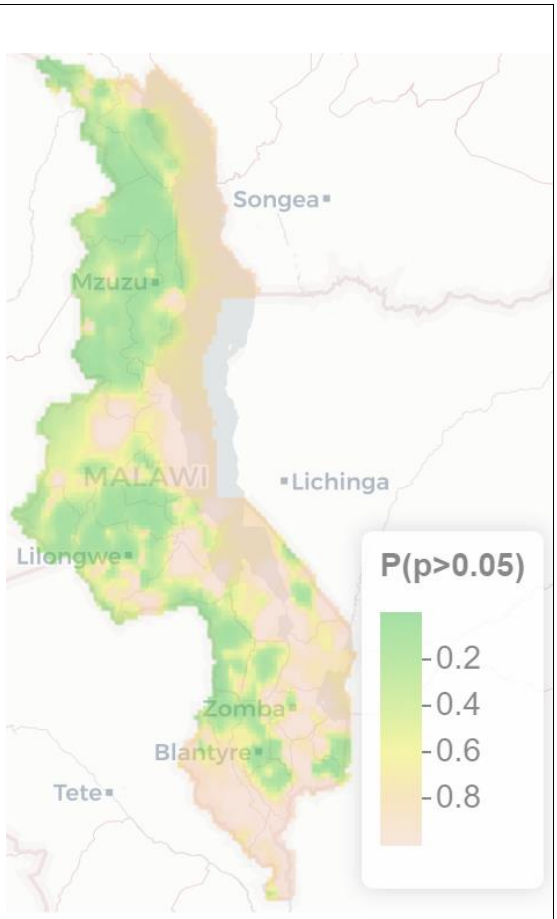


Figure 5.12 Exceedance probability of estimated prevalence for anySCHprev in Reassessment data

The four graphs below show the results of the Asc at the second survey. Some of the higher infection rates are still present in all locations. Compared to the first survey, the increase in infection rates is somewhat more pronounced in some areas. This indicates that the treatment did not control the disease very well.

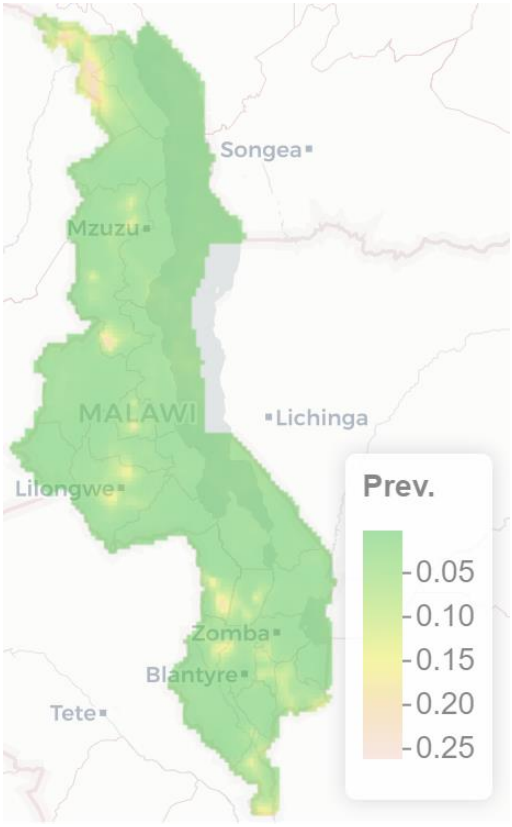


Figure 5.13 Map of estimated prevalence for Ascprev in Reassessment data

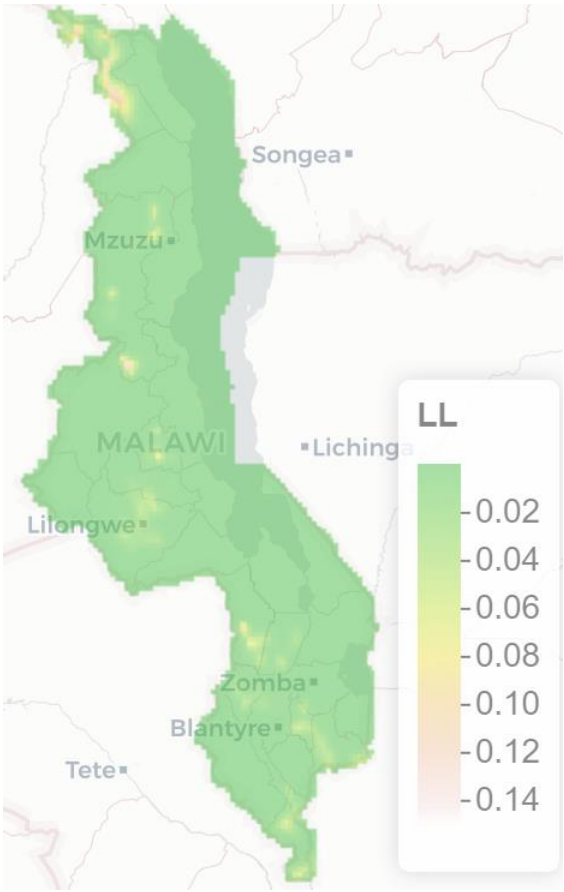


Figure 5.14 Lower limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Reassessment data

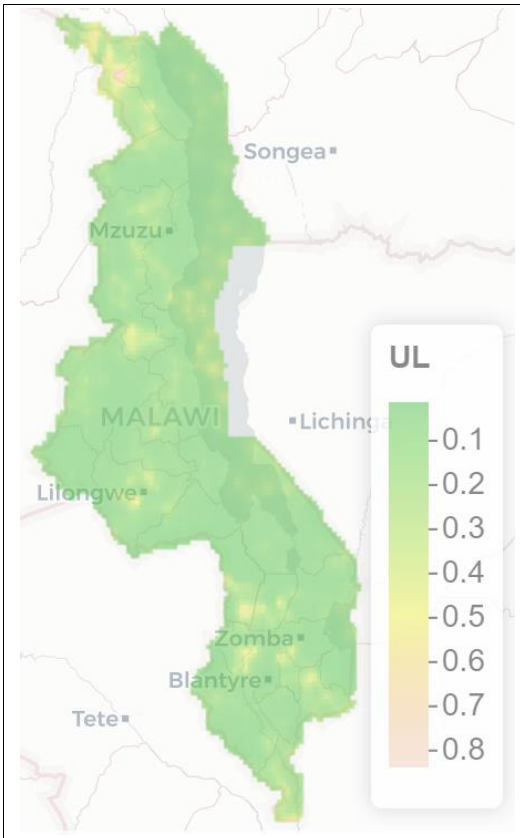


Figure 5.15 Upper limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Reassessment data

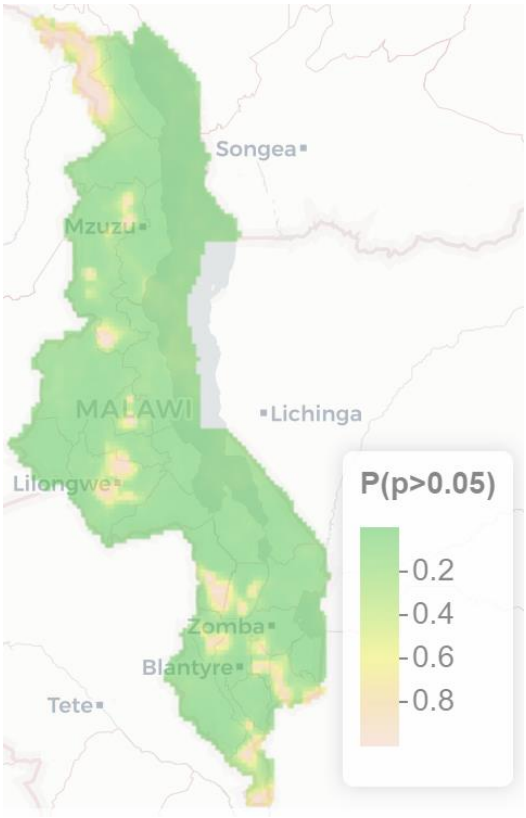


Figure 5.16 Exceedance probability of estimated prevalence for Ascprev in Reassessment data

The four graphs below show the results of Hkw at the time of the second survey. Some of the higher prevalence of infection is still present in all locations. Compared to the first survey, the increase in prevalence of infection is somewhat more pronounced in some areas. This indicates that the treatment did not control the disease very well.

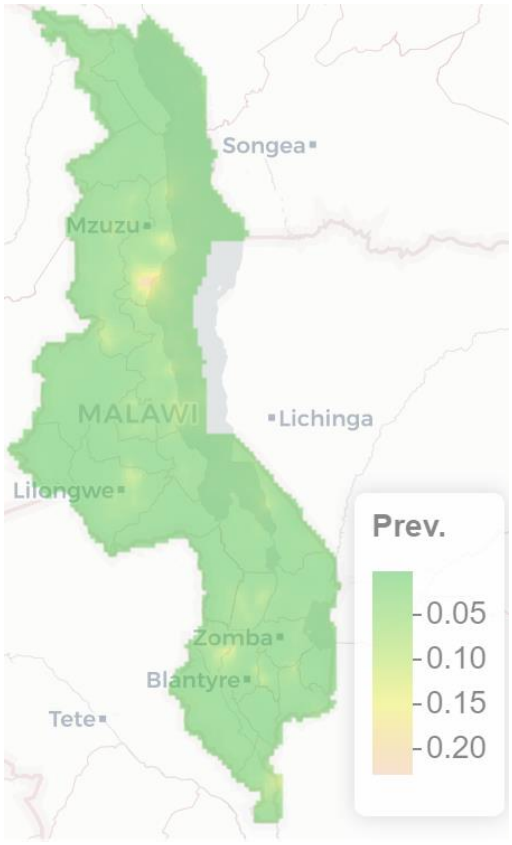


Figure 5.17 Map of estimated prevalence for Hkwprev in Reassessment data

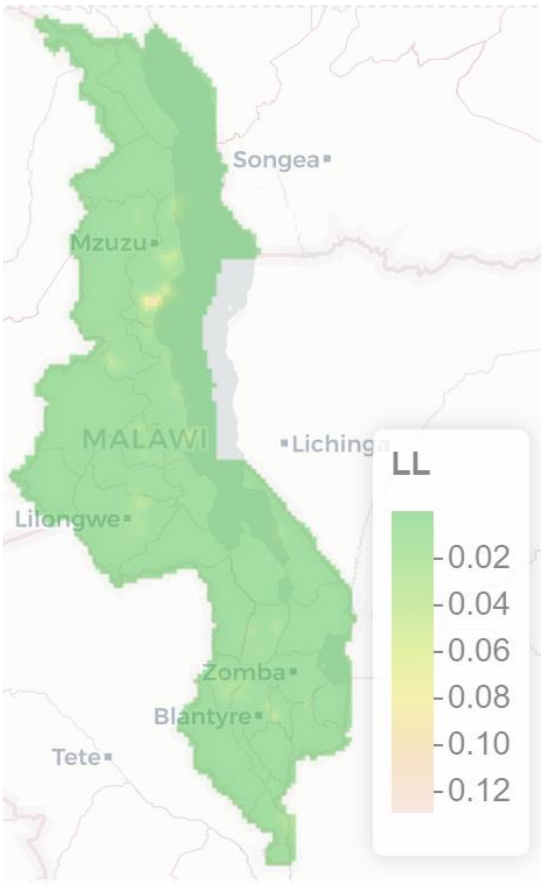


Figure 5.18 Lower limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Reassessment data

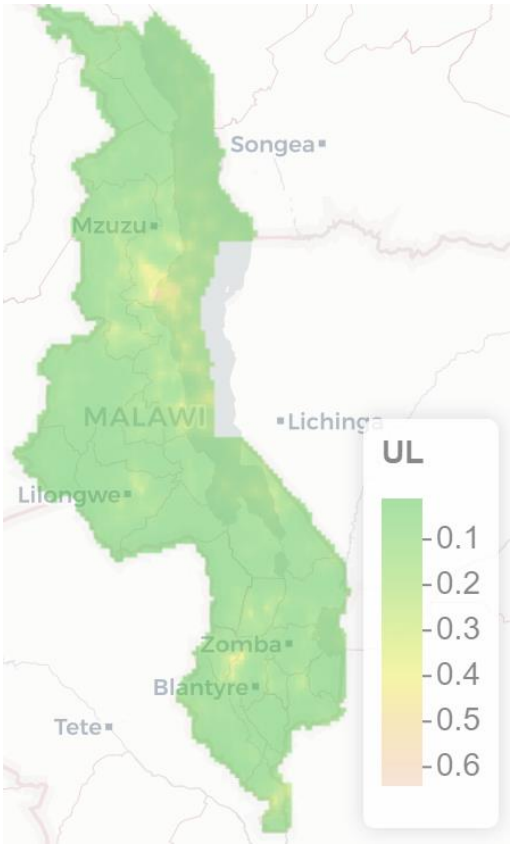


Figure 5.19 Upper limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Reassessment data

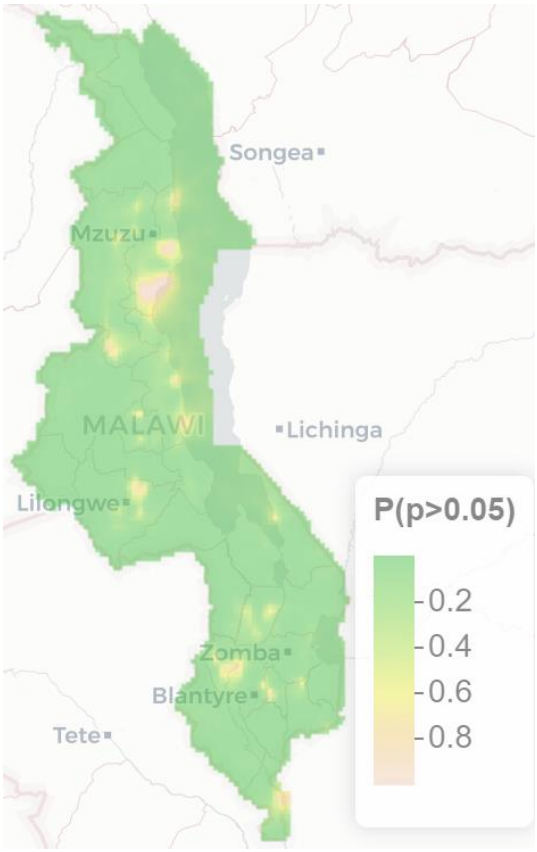


Figure 5.20 Exceedance probability of estimated prevalence for Hkwprev in Reassessment data

These four graphs show the results for Tri at the time of the second survey. the prevalence of infection in Tri has been very low and the lower 5% was chosen for the exceedance probability calculation, as none of the 20% calculations were met.

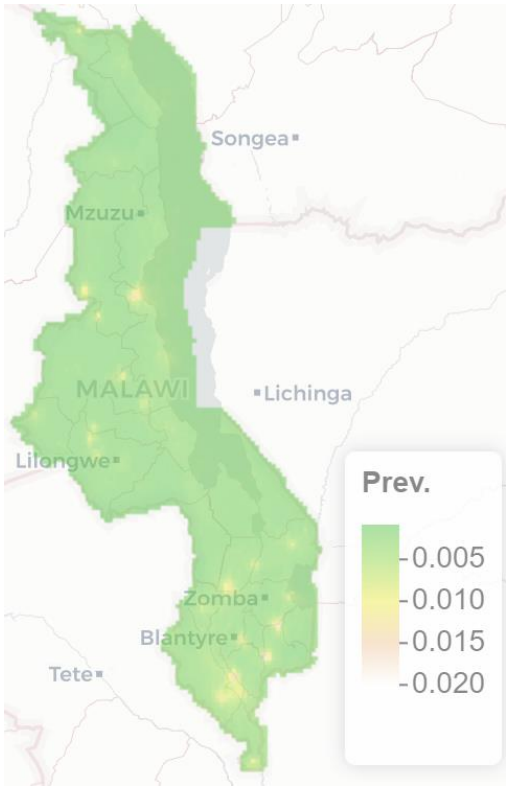


Figure 5.21 Map of estimated prevalence for Triprev in Reassessment data

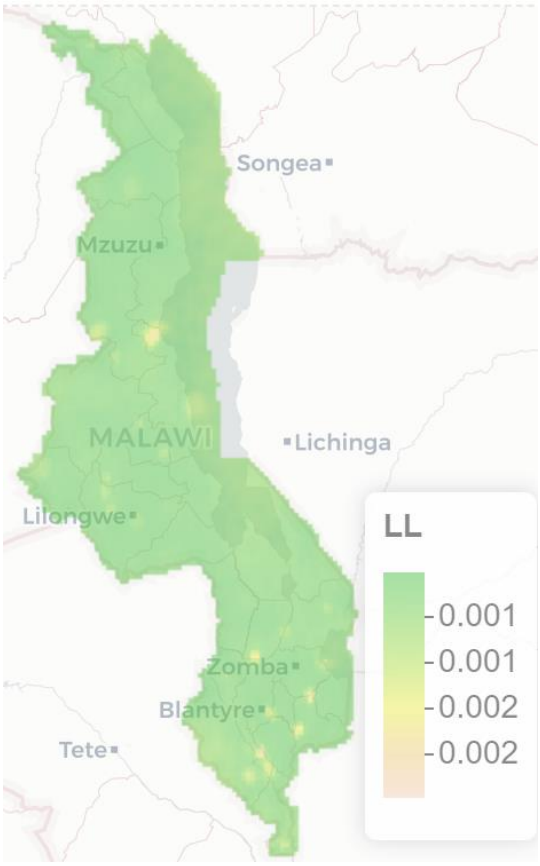


Figure 5.22 Lower limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Reassessment data

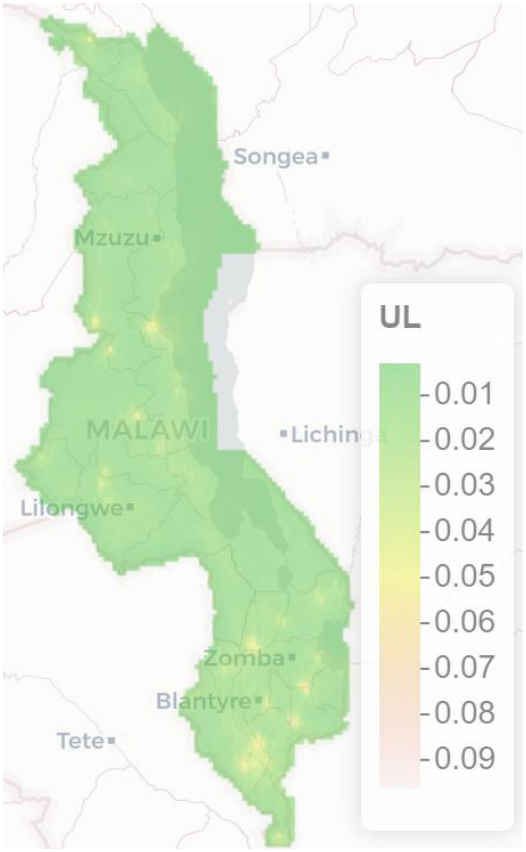


Figure 5.23 Upper limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Reassessment data

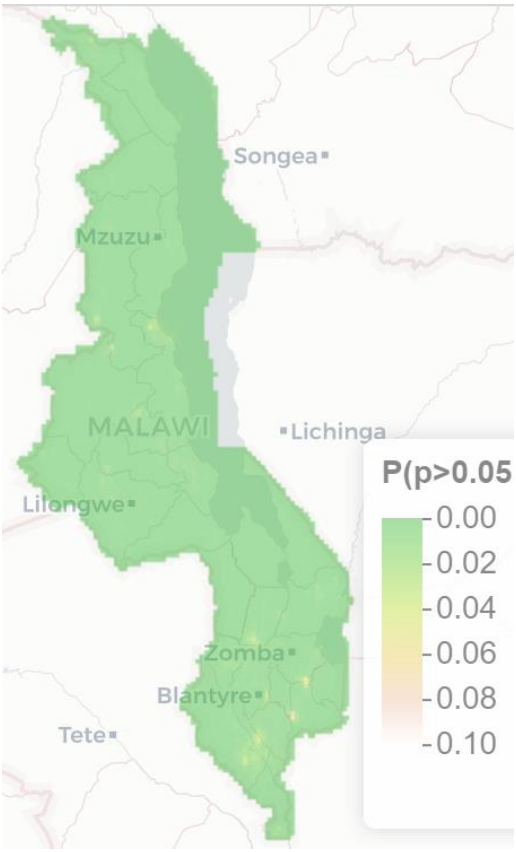


Figure 5.24 Exceedance probability of estimated prevalence for Triprev in Reassessment data

These four graphs show the sum of the results of the three STHs at the time of the second survey. Comparing the results of the first survey, STH has spread from being endemic only in the north to all parts of the country, which is not a good sign

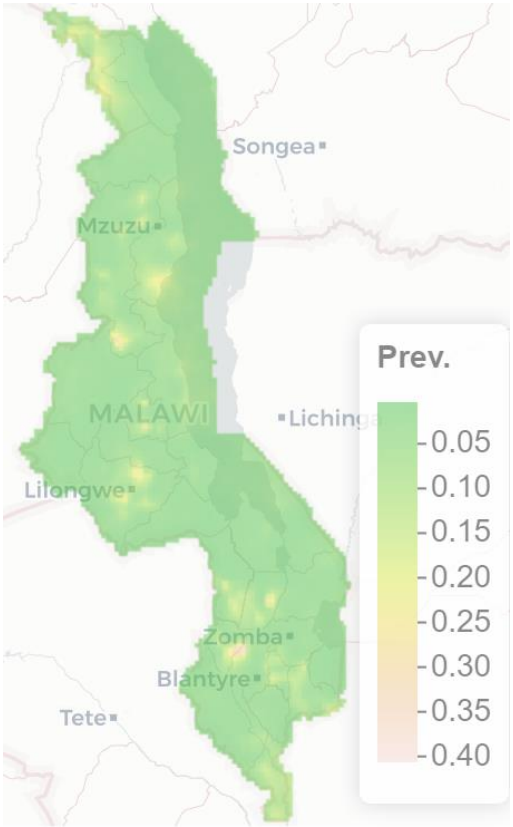


Figure 5.25 Map of estimated prevalence for anySTHprev in Reassessment data

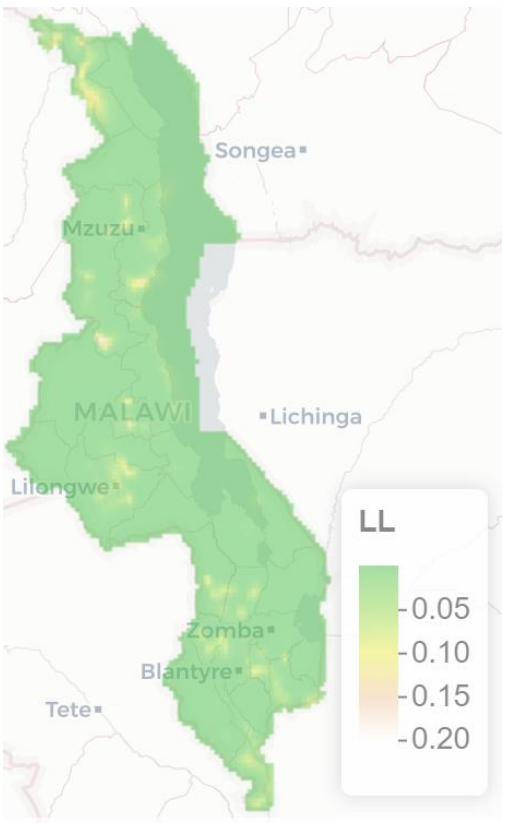


Figure 5.26 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Reassessment data

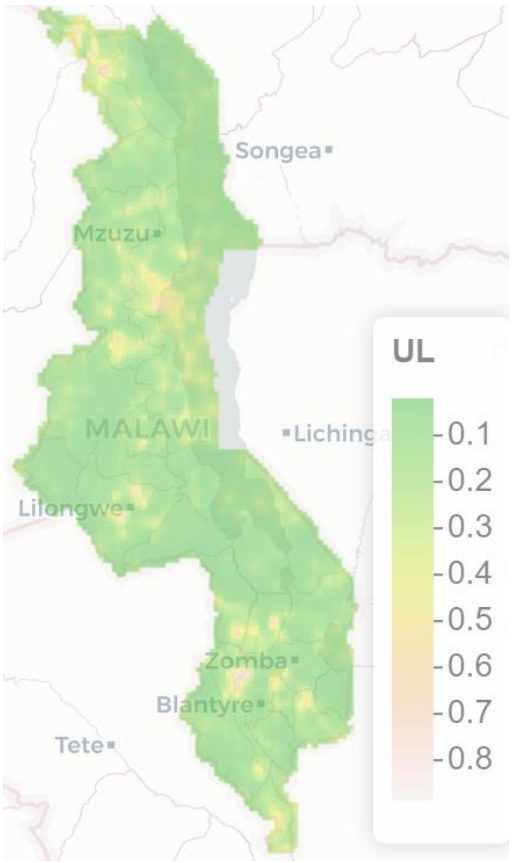


Figure 5.27 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Reassessment data

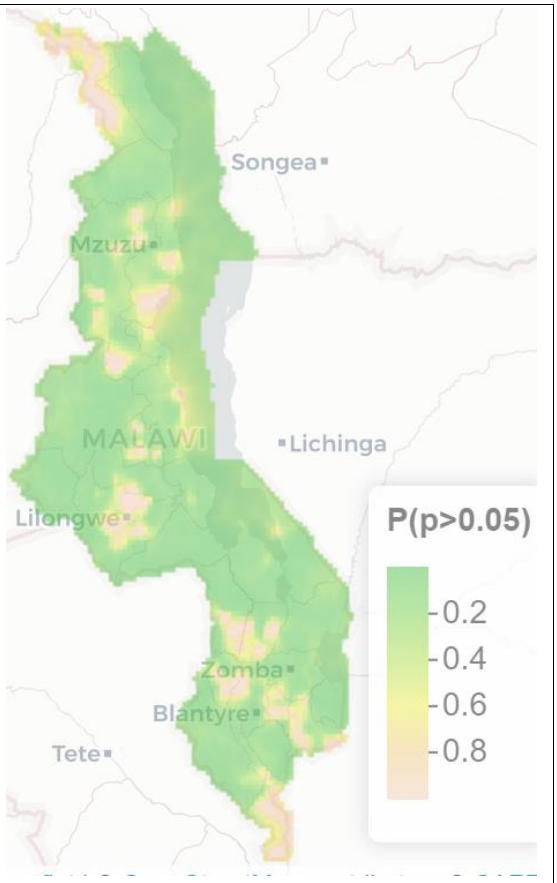


Figure 5.28 Exceedance probability of estimated prevalence for anySTHprev in Reassessment data

The following shows the prevalence of various diseases at the time of the first survey

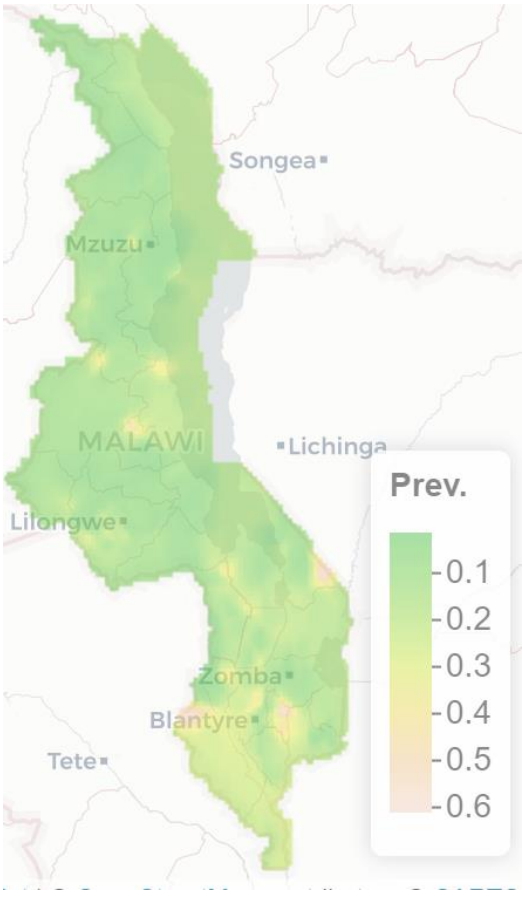


Figure 5.29 Map of estimated prevalence for Shprev in Historic data

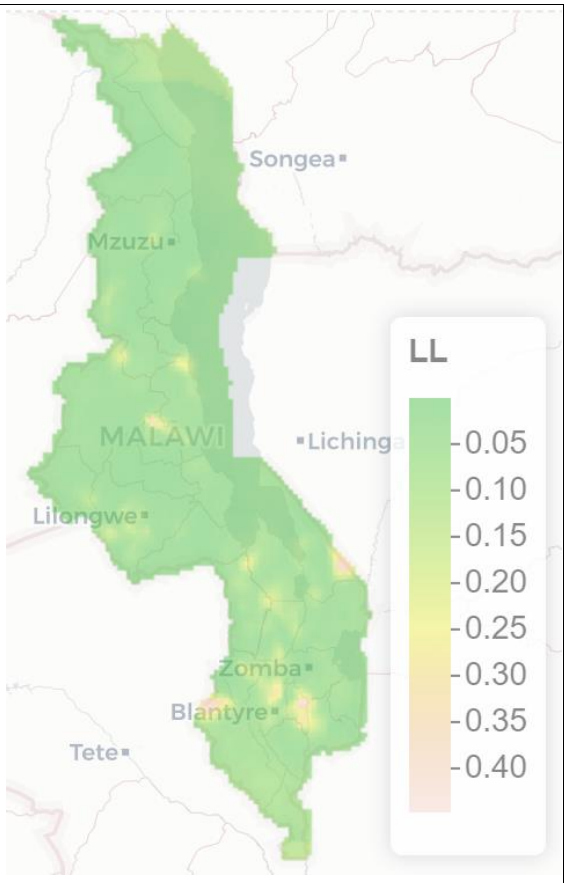


Figure 5.30 Lower limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Historic data

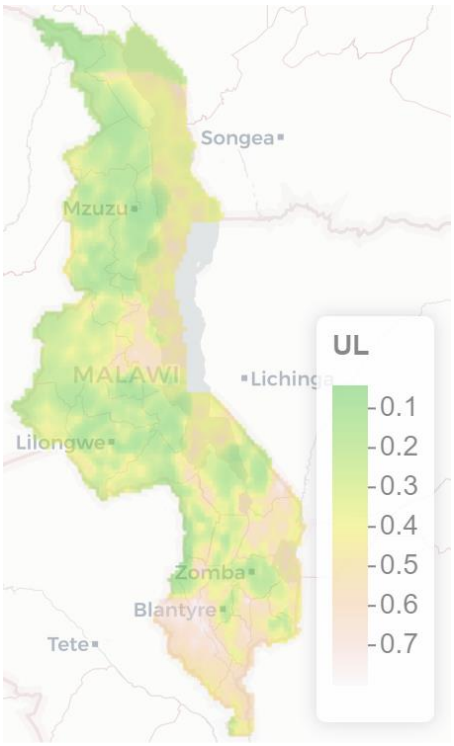


Figure 5.31 Upper limit of 95%

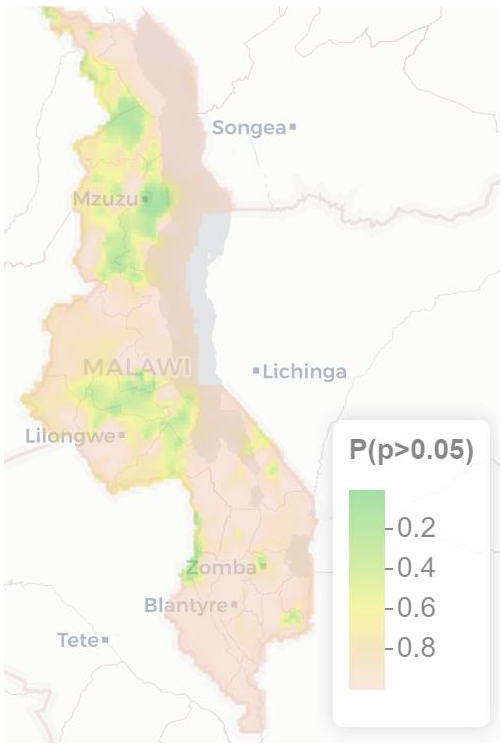


Figure 5.32 Exceedance probability of

confidence intervals for maps of estimated prevalence for Shprev in Historic data	estimated prevalence for Shprev in Historic data

Sman

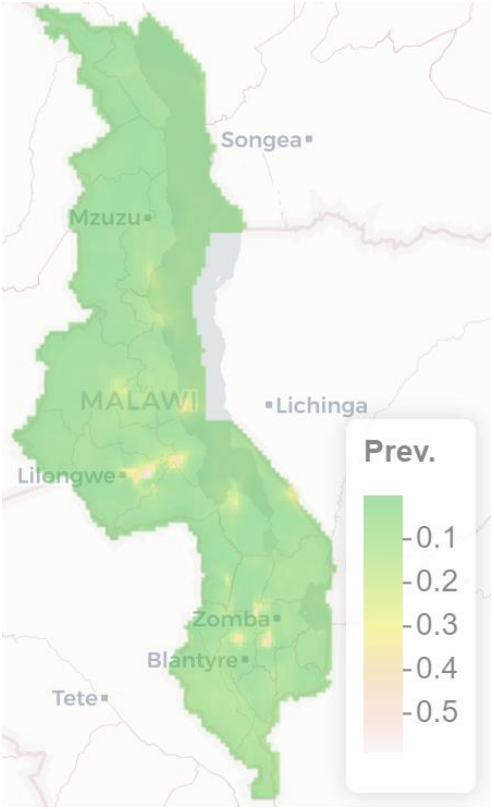


Figure 5.33 Map of estimated prevalence for Smanprev in Historic data

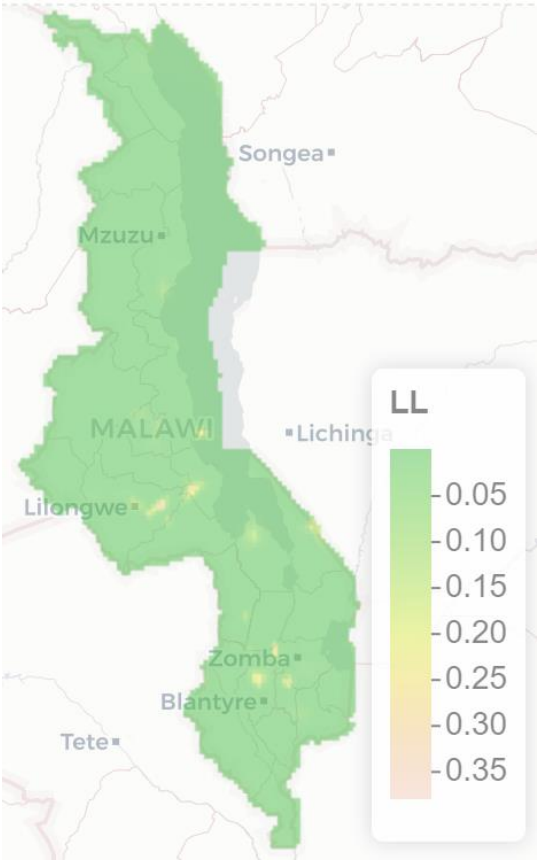


Figure 5.34 Lower limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Historic data

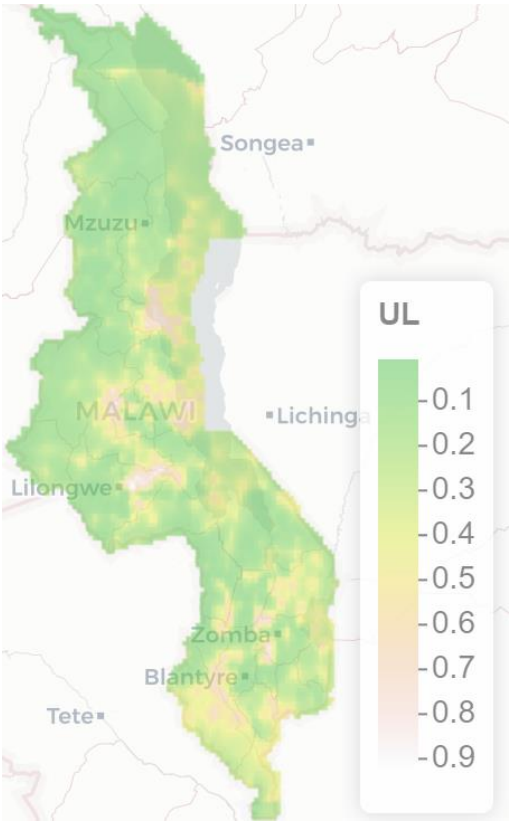


Figure 5.35 Upper limit of 95% confidence intervals for maps of estimated prevalence for Smanprev in Historic data

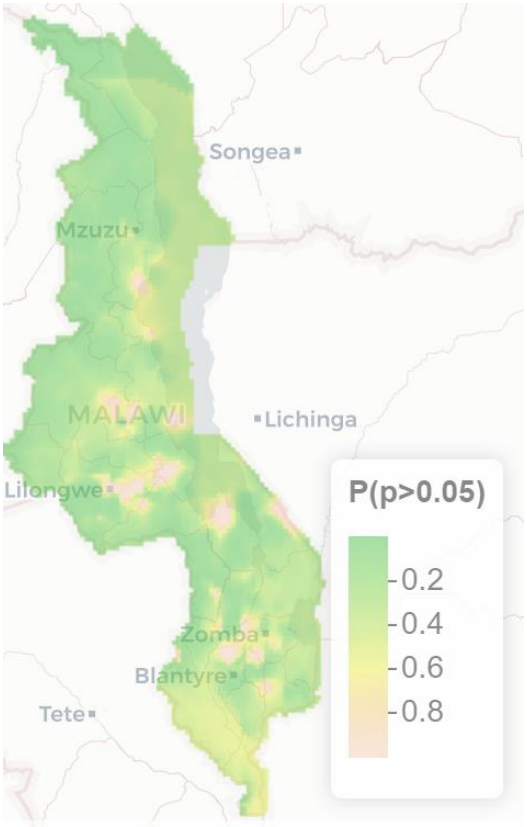


Figure 5.36 Exceedance probability of estimated prevalence for Smanprev in Historic data

SCH

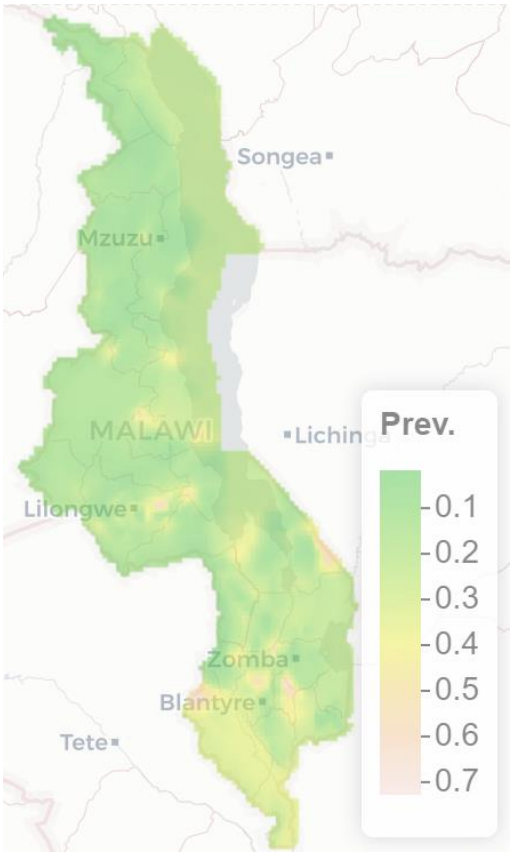


Figure 5.37 Map of estimated prevalence for anySCHprev in Historic data

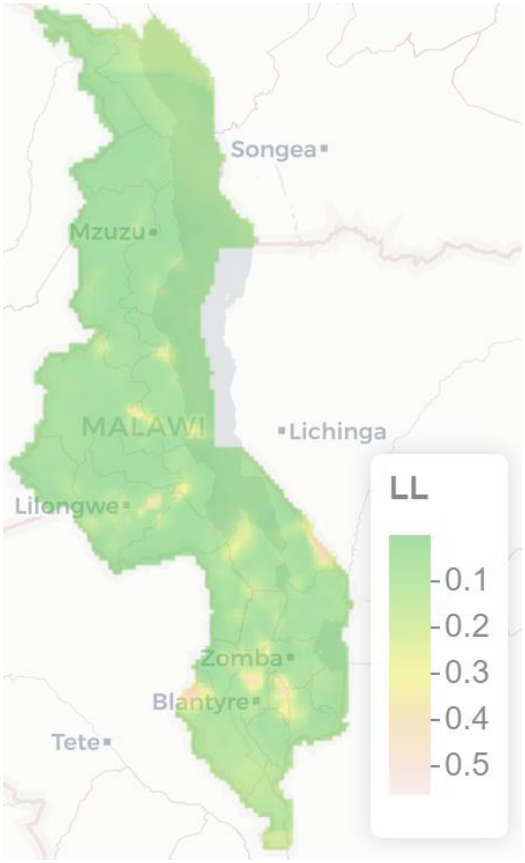


Figure 5.38 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Historic data

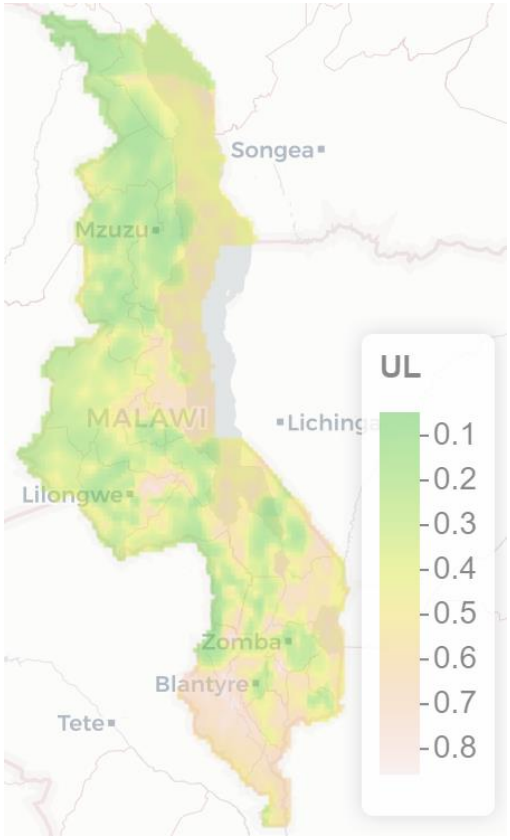


Figure 5.39 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySCHprev in Historic data

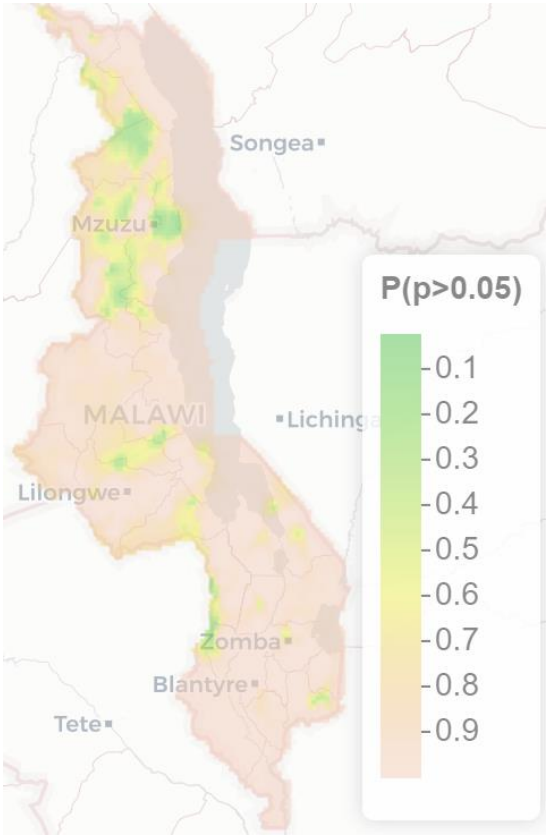


Figure 5.40 Exceedance probability of estimated prevalence for anySCHprev in Historic data

Asc

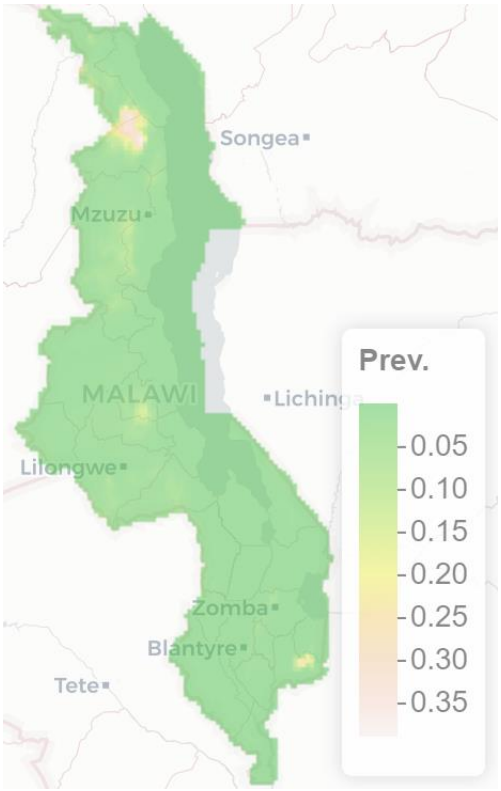


Figure 5.41 Map of estimated prevalence for Ascprev in Historic data

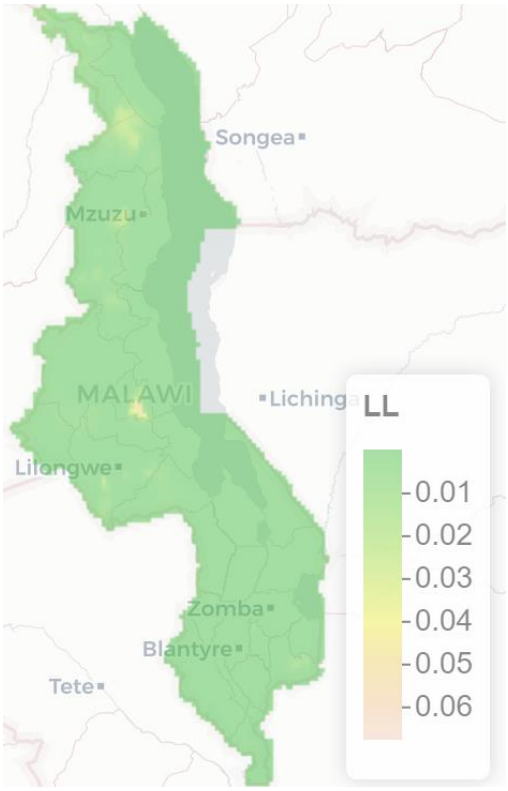


Figure 5.42 Lower limit of 95% confidence intervals for maps of estimated prevalence for Ascprev in Historic data

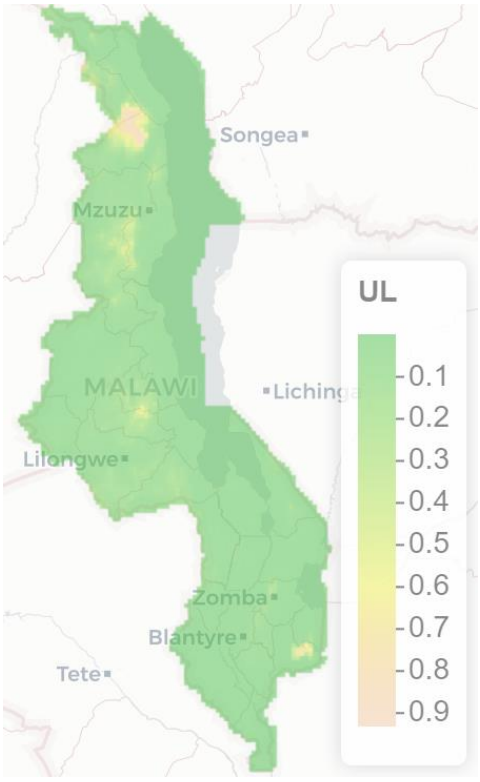


Figure 5.43 Upper limit of 95% confidence intervals for maps of

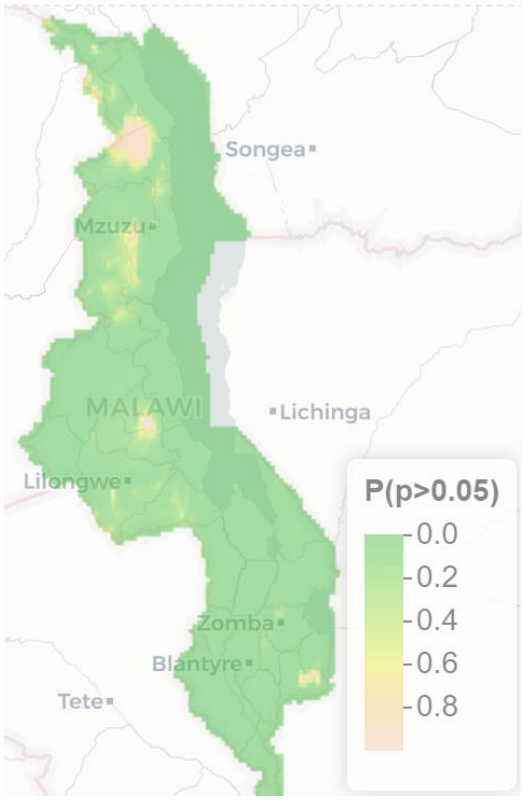


Figure 5.44 Exceedance probability of

estimated prevalence for Ascprev in Historic data	estimated prevalence for Ascprev in Historic data

Hkw

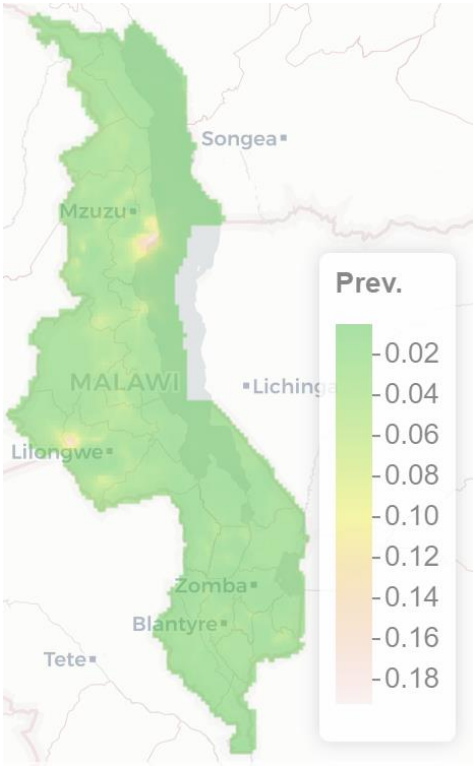


Figure 5.45 Map of estimated prevalence for Hkwprev in Historic data

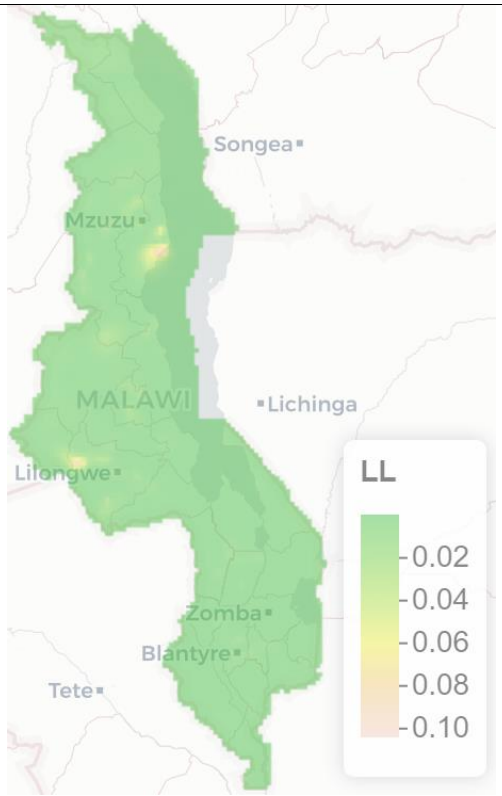


Figure 5.46 Lower limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Historic data

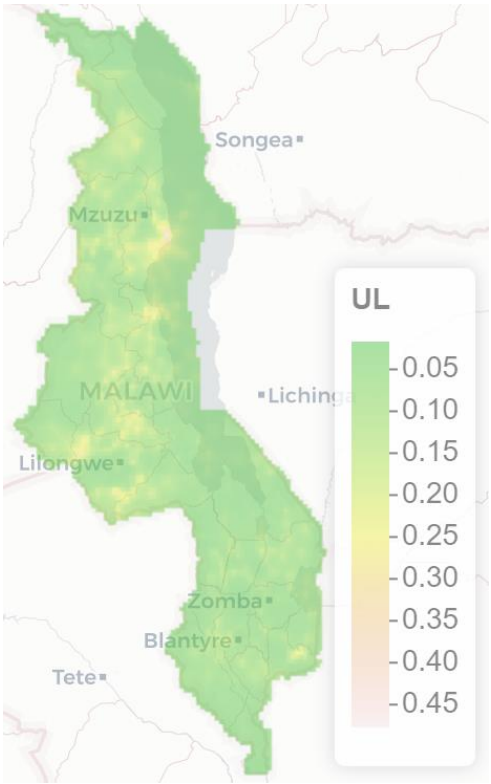


Figure 5.47 Upper limit of 95% confidence intervals for maps of estimated prevalence for Hkwprev in Historic data

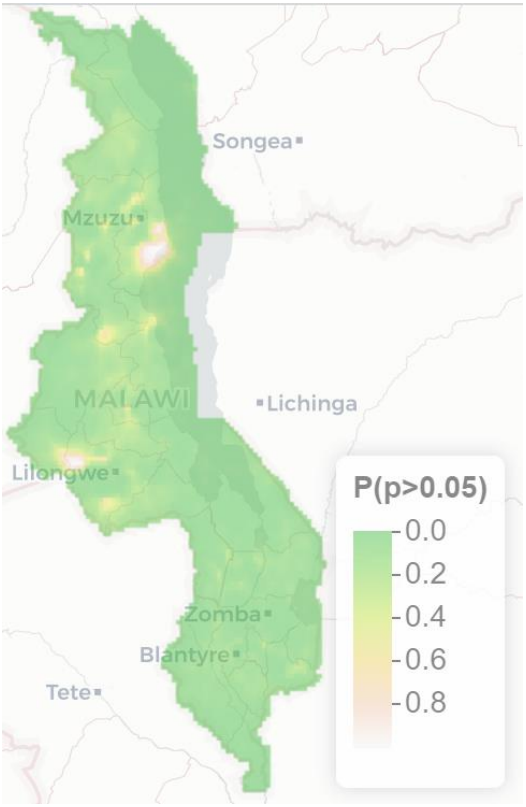


Figure 5.48 Exceedance probability of estimated prevalence for Hkwprev in Historic data

Tri

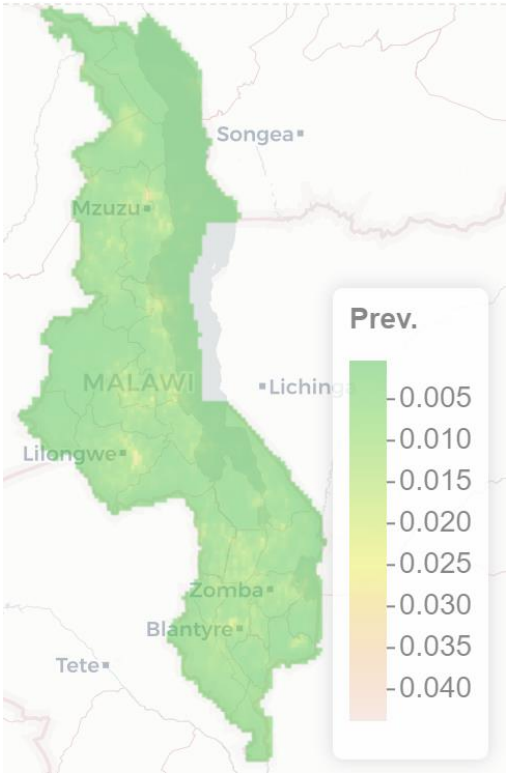


Figure 5.49 Map of estimated prevalence for Triprev in Historic data

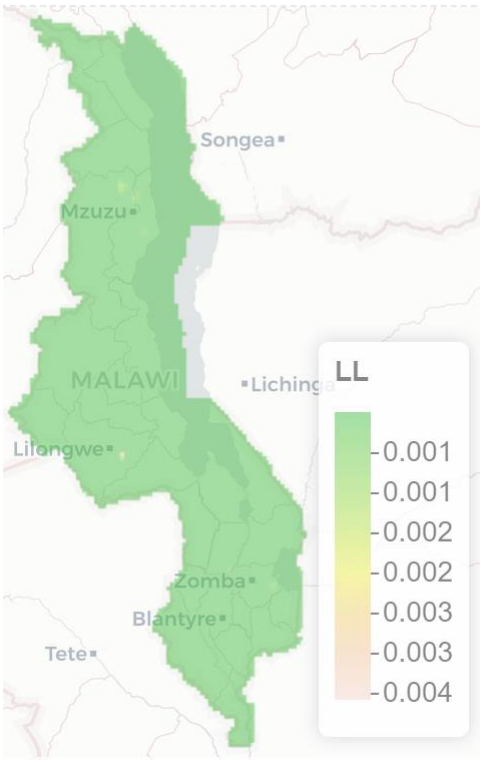
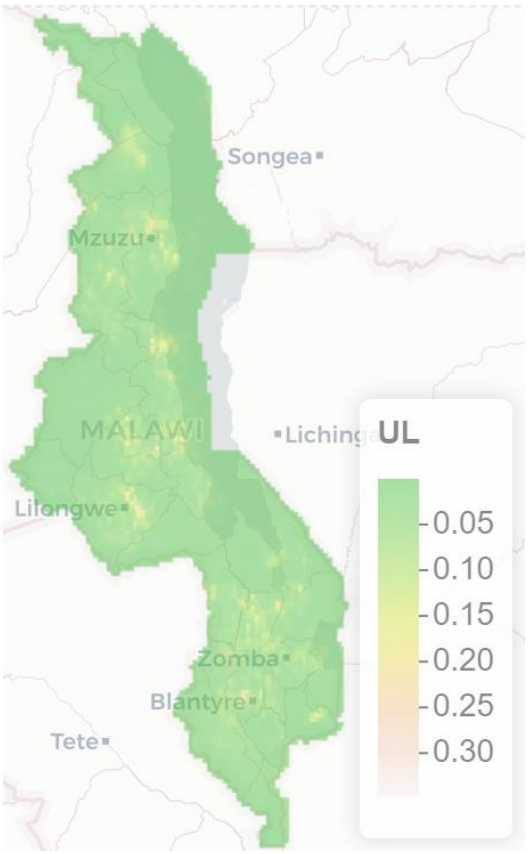


Figure 5.50 Lower limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Historic data



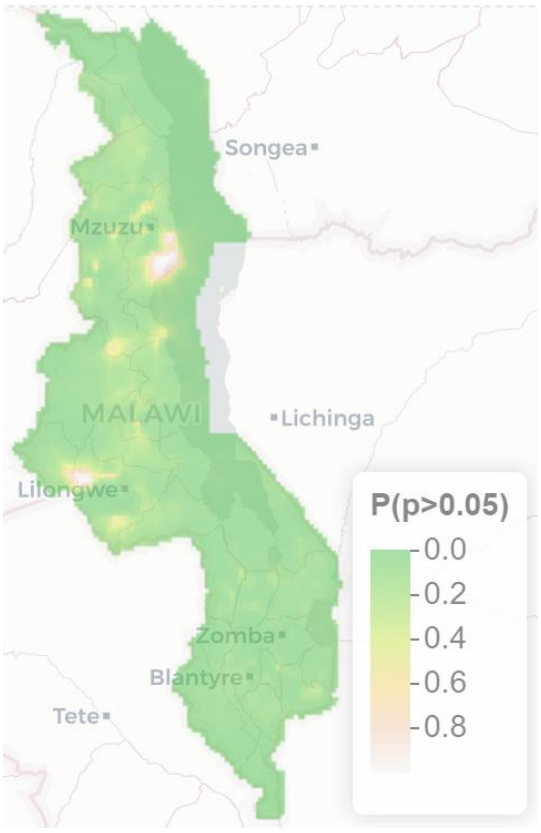


Figure 5.52 Exceedance probability of

Figure 5.51 Upper limit of 95% confidence intervals for maps of estimated prevalence for Triprev in Historic data	**estimated prevalence for Triprev in Historic data**

AnySTH

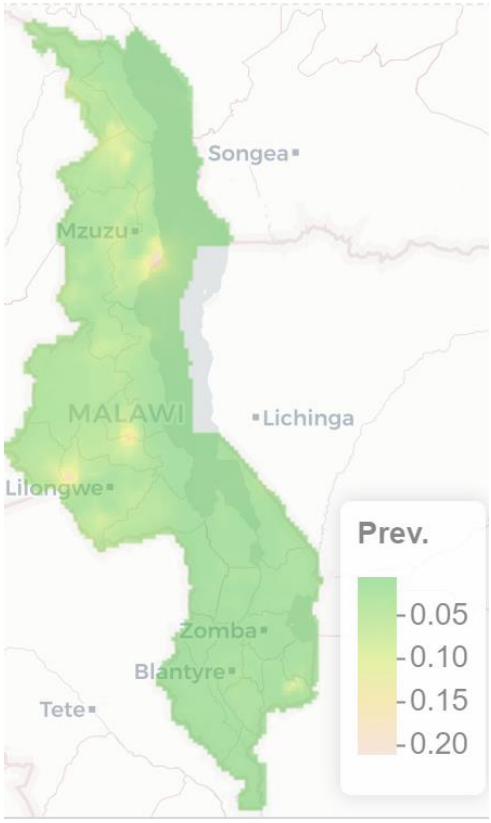


Figure 5.53 Map of estimated prevalence for anySTHprev in Historic data

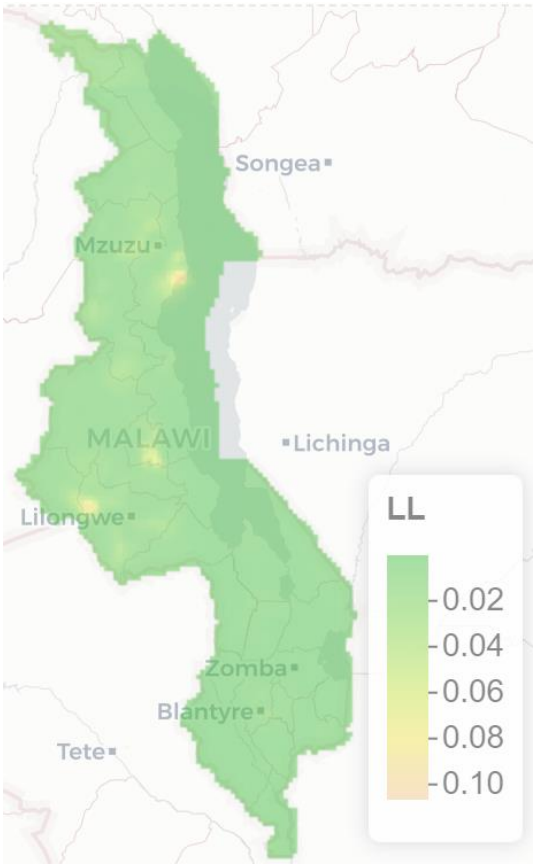


Figure 5.54 Lower limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Historic data

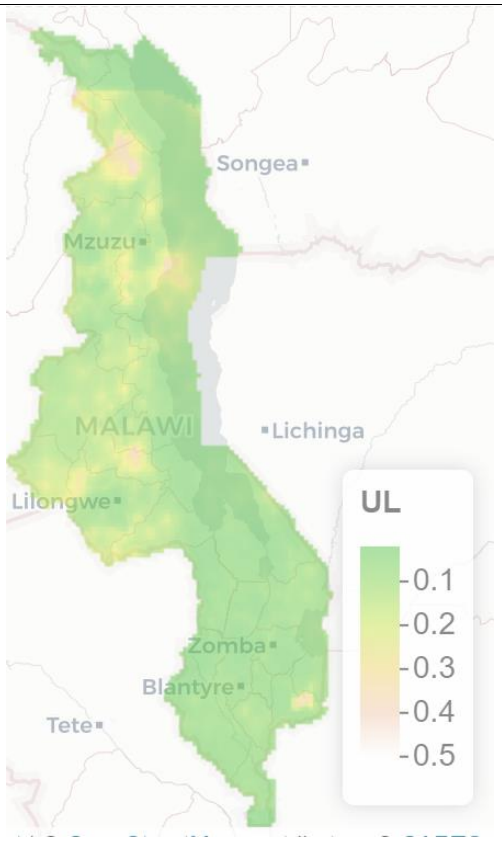


Figure 5.55 Upper limit of 95% confidence intervals for maps of estimated prevalence for anySTHprev in Historic data

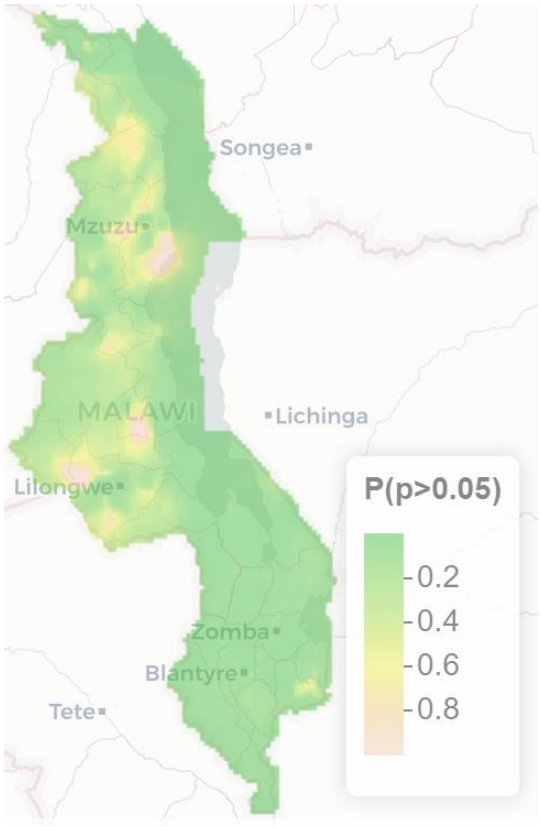


Figure 5.56 Exceedance probability of estimated prevalence for anySTHprev in Historic data

6. Discussion

This paper also attempts to estimate disease prevalence using the Prevmap package in R. Only the results of the first model were obtained because Prevmap is very slow and a full run of the model takes about an hour. This compares to INLA's speed of around half a minute. The pictures show the results of the second survey using Prevmap to estimate the infection rate of shprev.

It can be seen that the difference with the INLA results is small, but this approach was eventually abandoned due to the long computation time

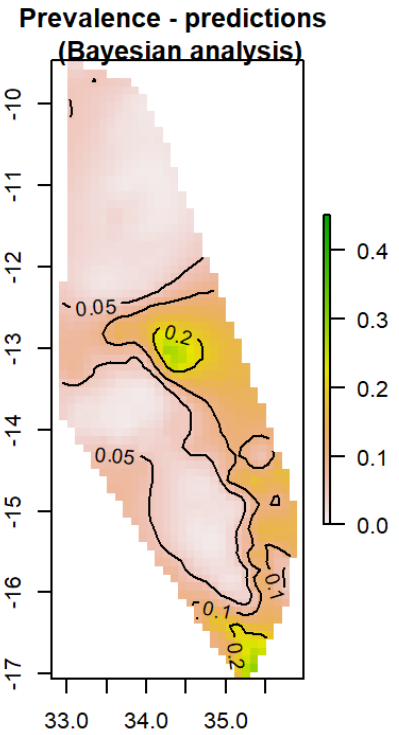


Figure 6.1 Map of estimated prevalence for Shprev in Reassessment data using Prevmap

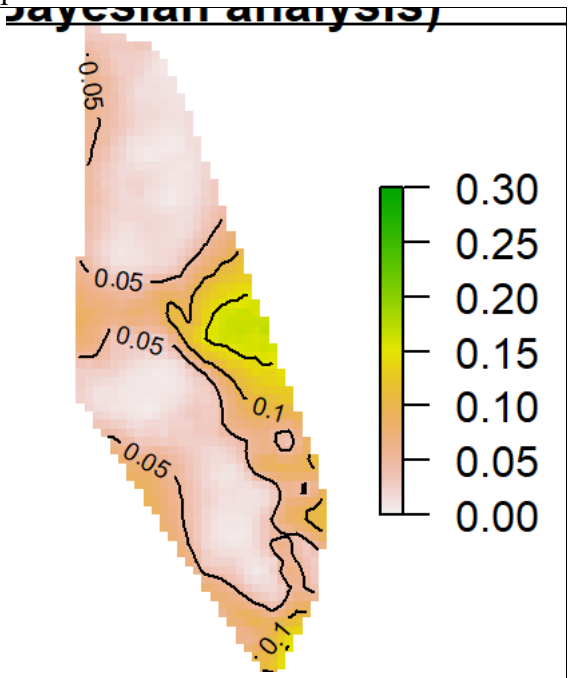


Figure 6.2 Standard Error for maps of estimated prevalence for Shprev in Reassessment data using Prevmap

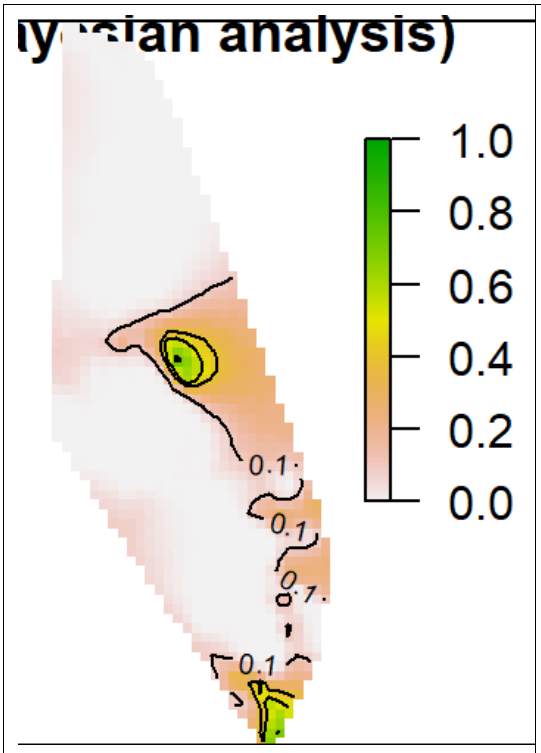


Figure 6.3 Exceedance probability of estimated prevalence for Shprev in Reassessment data using Prevmap

An attempt was also made to build a height-dependent spatial model, and it can be seen that the model results are very similar to the topographic heights in Malawi, and it is clear that the fit of such a model is also poor. The image shows the results of the infection rate estimates for shprev using the second INLA survey (only height was used as a variable)

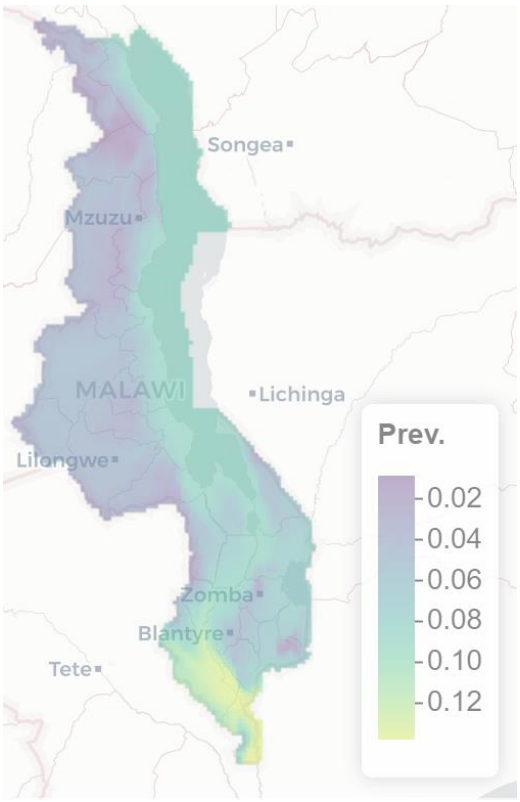


Figure 6.4 Map of estimated prevalence for Shprev in Reassessment data

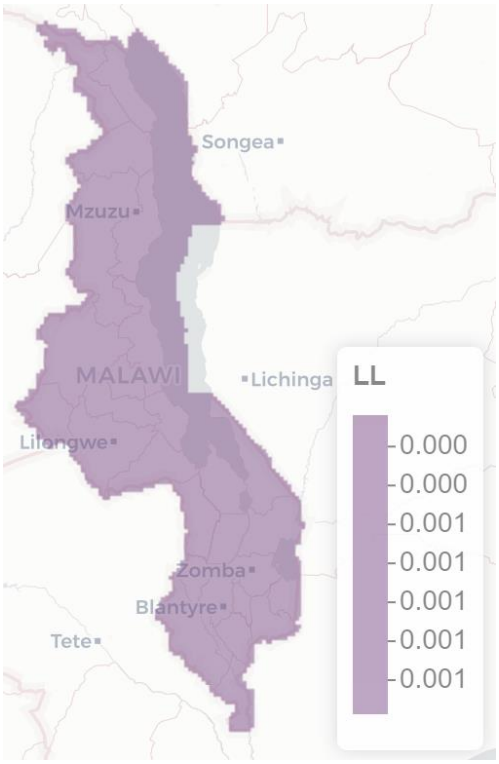


Figure 6.5 Lower limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data

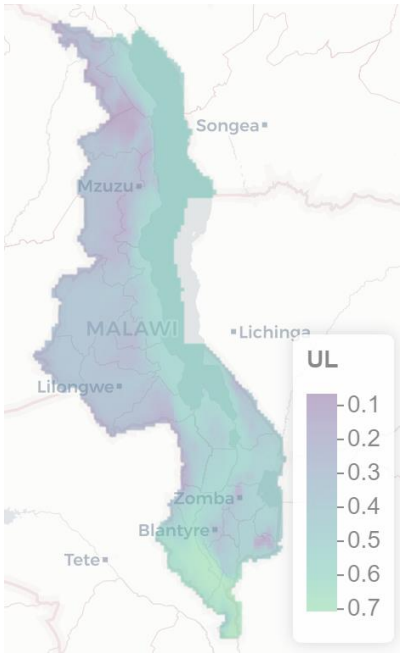


Figure 6.6 Upper limit of 95% confidence intervals for maps of estimated prevalence for Shprev in Reassessment data

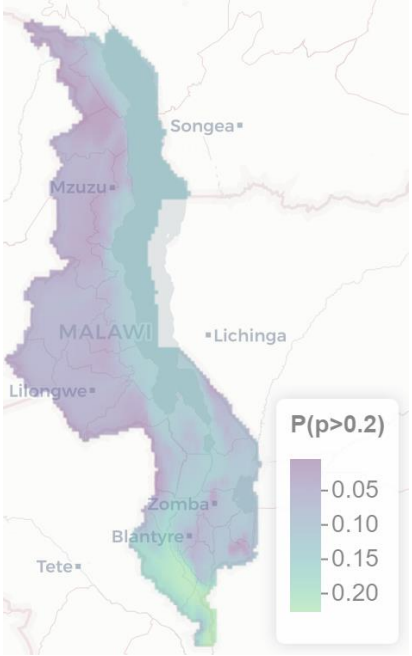


Figure 6.7 Exceedance probability of estimated prevalence for Shprev in Reassessment data

The final model in this paper has a number of areas for improvement, the biggest being the lack of time series analysis of the month data, the relationship with vegetation should also be considered and more data relating to disease should be added, the correlation between the available data and the results is not as clear as expected, and the lack of adequate validation of the model is also an area that needs to be improved in this paper. An attempt should have been made to obtain the results by different means and then compare them, but this was not achieved by the limited ability of the authors to code. For the future more could be done, if possible, to measure snail related intermediate hosts as well as water pollution, village sanitation levels etc. which would be of great help to the model, but unfortunately such data is difficult to find in Malawi. Malawi, a relatively poor and backward country, only completed its census in 2018. The authors often complain when modelling that the lack of good data makes the model not as highly desirable as one might think, with very close results after adding variables and not adding them, but this is probably a common problem in most backward developing countries, where poor medical care produces disease and underdeveloped levels of administration can make it difficult to cure. And the real problems are not like the examples in the books, where using the right method will give very desirable results. All we can do is to seek the truth and restore it. Thanks to the contribution of the WHO, schistosomiasis and soil-transmitted helminthiasis have improved in Malawi, but there is still a long way to go before they are completely eradicated in Malawi.

7. Conclusion

This paper maps the estimated prevalence of schistosomiasis and soil-transmitted helminthiasis in Malawi in 2012 and 2017 with a 20% or 5% probability of exceeding the prevalence estimates, and obtains 95% confidence intervals for the prevalence estimates. The study showed a large spatial correlation of diseases. The results also show that schistosomiasis and soil-transmitted helminthiasis are still some time away from being eradicated.

Reference

[1] Grimes, J., Croll, D., Harrison, W., Utzinger, J., Freeman, M. and Templeton, M., 2015. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. Parasites & Vectors, 8(1).
[2] Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. Lancet Infect Dis. 2006;6:411–25.
[3] Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. Lancet. 2006;368:1106–18.
[4] Colley DG, Bustinduy AL, Secor WE, King CH. Human schistosomiasis. Lancet. 2014;383:2253–64.
[5] Rollinson D. A wake up call for urinary schistosomiasis: reconciling research effort with public health importance. Parasitology. 2009;136:1593–610.
[6] Secor WE. The effects of schistosomiasis on HIV/AIDS infection, progression and transmission. Curr Opin HIV AIDS. 2012;7:254–9.
[7] Okwuosa VN, Osuala FO. Toxicity of washing soaps to Schistosoma mansoni cercariae and effects of sublethal concentrations on infectivity in mice. Appl Parasitol. 1993;34:69–75.
[8] Birrie H, Balcha F, Erko B, Bezuneh A, Gemeda N. Investigation into the cercariacidal and miracidiacidal properties of endod (Phytolacca dodecandra) berries (type 44). East Afr Med J. 1998;75:311–4.

[9] Berke O . Choropleth mapping of regional count data of Echinococcus multilocularis among red foxes in Lower Saxony，Germany . Preventive Veterinary Medicine，2001，52(2) : 119-131 .

[10] Core.ac.uk. 2022. 空间流行病学中的疾病制图常用方法. [online] Available at: <https://core.ac.uk/reader/41438067> [Accessed 9 September 2022].
[11] Jianshu.com. 2022. [online] Available at: <https://www.jianshu.com/p/d0566b655db9> [Accessed 9 September 2022].
[12] Moraga, P., 2022. Welcome | Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. [online] Paulamoraga.com. Available at: <https://www.paulamoraga.com/book-geospatial/> [Accessed 9 September 2022].

[13] Sidén, P. and Lindsten, F., 2022. Deep Gaussian Markov Random Fields. [online] arXiv.org. Available at: <https://arxiv.org/abs/2002.07467> [Accessed 9 September 2022].

[14] Blog.csdn.net. 2022. 高斯马尔科夫随机场_weixin_33984032 的博客-CSDN 博

客 . [online] Available at: <https://blog.csdn.net/weixin_33984032/article/details/94523625> [Accessed 9 September 2022].
[15] Punam Amratia, A., 2022. Introduction to INLA for geospatial modelling. [online] Punama.github.io. Available at: <https://punama.github.io/BDI_INLA/> [Accessed 9 September 2022].
[16] Makaula, P., Sadalaki, J., Muula, A., Kayuni, S., Jemu, S. and Bloch, P., 2014. Schistosomiasis in Malawi: a systematic review. Parasites & Vectors, 7(1).
[17] Liao, C., Fu, C., Kao, C., Lee, Y., Chen, P., Chuang, T., Naito, T., Chou, C., Huang, Y., Bonfim, I. and Fan, C., 2016. Prevalence of intestinal parasitic infections among school children in capital areas of the Democratic Republic of São Tomé and Príncipe, West Africa. African Health Sciences, 16(3), p.690.
[18] Lai, Y., Biedermann, P., Ekpo, U., Garba, A., Mathieu, E., Midzi, N., Mwinzi, P., N'Goran, E., Raso, G., Assaré, R., Sacko, M., Schur, N., Talla, I., Tchuenté, L., Touré, S., Winkler, M., Utzinger, J. and Vounatsou, P., 2015. Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. The Lancet Infectious Diseases, 15(8), pp.927-940.
[19] Hodges, M., Soares Magalhães, R., Paye, J., Koroma, J., Sonnie, M., Clements, A. and Zhang, Y., 2012. Combined Spatial Prediction of Schistosomiasis and Soil-Transmitted Helminthiasis in Sierra Leone: A Tool for Integrated Disease Control. PLoS Neglected Tropical Diseases, 6(6), p.e1694.
[20] Ekpo, U., Hürlimann, E., Schur, N., Oluwole, A., Abe, E., Mafe, M., Nebe, O., Isiyaku, S., Olamiju, F., Kadiri, M., Poopola, T., Braide, E., Saka, Y., Mafiana, C., Kristensen, T., Utzinger, J. and Vounatsou, P., 2013. Mapping and prediction of schistosomiasis in Nigeria using compiled survey data and Bayesian geospatial modelling. Geospatial health, 7(2), p.355.
[21] Schur, N., Hürlimann, E., Garba, A., Traoré, M., Ndir, O., Ratard, R., Tchuem Tchuenté, L., Kristensen, T., Utzinger, J. and Vounatsou, P., 2011. Geostatistical Model-Based Estimates of Schistosomiasis Prevalence among Individuals Aged ⩽20 Years in West Africa. PLoS Neglected Tropical Diseases, 5(6), p.e1194.
[22] RASO, G., MATTHYS, B., N'GORAN, E., TANNER, M., VOUNATSOU, P. and UTZINGER, J., 2005. Spatial risk prediction and mapping of Schistosoma mansoni infections among schoolchildren living in western Côte d'Ivoire. Parasitology, 131(1), pp.97-108.

Appendix

Sub-district schistosomiasis and STH endemicity classification in Malawi

Ruotong Xiao
Project hosted by SCI Foundation

Motivation

The Malawian MoH has an on-going national scale program for schistosomiasis and STH in all of its 29 districts. SCIF supported baseline mapping (2012-13) and all districts were endemic for both Schistosoma mansoni and S. haematobium. National scale annual treatment with praziquantel (PZQ) for schistosomiasis and albendazole (ALB) for STH commenced in 2012. According to WHO guidelines, after five to six rounds of treatment, an evaluation should
be conducted (WHO, 2011) to re-evaluate treatment frequency and ensure the program focuses resources where the need is greatest. SCIF supported the national reassessment survey between 2017-2019. In both the original mapping survey and the reassessment, the evaluation area was the district.
The drugs required for national treatment campaigns are donated through the WHO's drug donation program. In support of the new WHO guidelines for schistosomiasis control and elimination, the African Regional Office for WHO require countries to submit their drug requests based on sub-district administrative units. Few countries have sufficient infection data at this sub-district level to make accurate prevalence estimates needed for program decision-making.
This project will support the Ministry of Health (MoH) in Malawi to update their national strategy for schistosomiasis and soiltransmitted helminthiasis (STH) in line with the WHO guidelines and enable sub-district implementation.

Data

The data is divided into three parts, firstly a map of the sub-district administrative unit level (Traditional Authority level), then two files of infection rates before and after drug treatment The files are small and the data is relatively simple

Aims

The project will apply model-based geostatistical techniques to existing epidemiological data to produce sub-district endemicity classification (and associated uncertainty) of schistosomiasis and STH.
[Possible extension – calculate sub-district level population estimates using publicly available population density data]

Timeline:

July:
Fully understand the task, read the literature, study the geography module and build a model

August:
Make multiple models, perform visual analysis from using multiple models

Changes may be made to the aims and objective during the project, depending on the success, challenges, and time scale.

Deliverables

A scoping document to agree approach (e.g., exceedance probability, prevalence estimates etc.), which sets out the key questions that the modelling will help to answer
- Report including:
o Summary of methods used
o tables of Traditional Authority (TA) level endemicity classification of schistosomiasis and STH from 2017-2019 reassessment data.
o Tables of TA level endemicity classification of schistosomiasis and STH from 2012-13 national reassessment data
o Prevalence risk surface maps
o Uncertainty associated with endemicity classification
o Other appropriate data visualization/tabulations