Heliyon 9 (2023) e20695

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Model-based geostatistical design and analysis of prevalence for soil-transmitted helminths in Kenya: Results from ten-years of the Kenya national school-based deworming programme

Collins Okoyo [a,b,c,*], Mark Minnery [d], Idah Orowe [a], Chrispin Owaga [e], Suzy J. Campbell [d], Christin Wambugu [f], Nereah Olick [g], Jane Hagemann [d], Wyckliff P. Omondi [h], Kate McCracken [d], Antonio Montresor [i], Graham F. Medley [j], Claudio Fronterre [k], Peter Diggle [k], Charles Mwandawiro [b]

[a] School of Mathematics, University of Nairobi, Nairobi, Kenya
[b] Eastern and Southern Africa Centre of International Parasite Control (ESACIPAC), Kenya Medical Research Institute, Nairobi, Kenya
[c] Department of Epidemiology, Statistics and Informatics (DESI), Kenya Medical Research Institute, Nairobi, Kenya
[d] Deworm the World, Evidence Action, Washington DC, United States
[e] Deworm the World, Evidence Action, Nairobi, Kenya
[f] Division of Adolescent and School Health (DASH), Ministry of Health, Nairobi, Kenya
[g] School Health, Nutrition and Meals Unit (SHNMU), Ministry of Education, Nairobi, Kenya
[h] Division of Vector Borne and Neglected Tropical Diseases (DVBNTD), Ministry of Health, Nairobi, Kenya
[i] Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland
[j] Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom
[k] Centre for Health Informatics, Computing and Statistics (CHICAS), Lancaster Medical School, Lancaster University, Lancaster, United Kingdom

ARTICLE INFO

Keywords:
Soil-transmitted helminths
Geostatistics
Model-based geostatistics
Modelling
National school-based deworming
Prevalence
Kenya

ABSTRACT

Background: Kenya is endemic for soil-transmitted helminths (STH) with over 6 million children in 27 counties currently at-risk. A national school-based deworming programme (NSBDP) was launched in 2012 with a goal to eliminate parasitic worms as a public health problem. This study used model-based geostatistical (MBG) approach to design and analyse the impact of the NSBDP and inform treatment strategy changes.

Methods: A cross-sectional study was used to survey 200 schools across 27 counties in Kenya. The study design, school selection and analysis followed the MBG approach which incorporated historical data on treatment, morbidity and environmental covariates to efficiently predict the helminths prevalence in Kenya.

Results: Overall, the NSBDP geographic area prevalence for any STH was estimated to sit between 2 % and <10 % with a high predictive probability of >0.999. Species-specific thresholds were between 2 % and <10 % for *Ascaris lumbricoides,* 0 % to <2 % for hookworm, and 0 % to <2 % for *Trichuris trichiura*, all with high predictive probability of >0.999.

Conclusions: Based on the World Health Organization guidelines, STH treatment requirements can now be confidently refined. Ten counties may consider suspending treatment and implement appropriate surveillance system, while another 10 will require treatment once every two years, and the remaining seven will require treatment once every year.

* Corresponding author. Eastern and Southern Africa Centre of International Parasite Control (ESACIPAC), Kenya Medical Research Institute, P.O. Box 54840-00200, Nairobi, Kenya.
E-mail address: comondi@kemri.go.ke (C. Okoyo).

https://doi.org/10.1016/j.heliyon.2023.e20695
Received 2 June 2023; Received in revised form 2 October 2023; Accepted 4 October 2023
Available online 5 October 2023

1. Introduction

1.1. Global and Kenya worm burden

Soil-transmitted helminths (STH) are among the most wide-spread neglected tropical diseases (NTDs) globally. STH affects more than 1.5 billion people of the world's poorest population [1]. Children of school-age are particularly vulnerable to chronic infection. This can impair their mental and physical development, reduce school attendance and educational achievement [2]. Infection with STH can lead to local and systemic pathological effects including anaemia, growth stunting, impaired cognition, decreased physical fitness and organ-specific effects. Severe cases can lead to intestinal obstructions and gangrene [3]. Kenya is endemic with both STH and schistosomiasis, with over 6 million children at-risk of parasitic worm infection [4].

1.2. Interventions for reducing worm burden

Repeated preventative chemotherapy (PC) with albendazole or mebendazole is used to control helminth morbidity within at-risk populations [5]. The two drugs are well suited for PC given their known safety profile, tolerability and low cost [6], and are often administered through school-based deworming campaigns [7]. Using schools as a platform for PC allows a "captive" population for treatment, maintaining high levels of coverage while minimizing cost and targeting those at most risk [8]. Regular school-based deworming is a proven and cost-effective strategy that can avert the health and educational consequences of STH [9].

1.3. Mapping and impact surveys for deworming programme

Kenya launched the national school-based deworming programme (NSBDP), in 2012. The goal of the programme was to eliminate parasitic worms as a public health problem (EPHP) in Kenya, by providing PC to all school-age children (SAC) and preschool-age children (PSAC) in selected counties across the country [7]. The NSBDP has an annual targeted to deworm over 6 million children across 27 STH endemic counties in parts of seven regions. In this way, it aims to treat over 80 % of all PSAC and SAC aged between 2 and 14 years for STH, within all endemic areas. Treatment was based on World Health Organization (WHO) guidelines as determined at the beginning of the programme.

The STH prevalence and the NSBDP assessment surveys have been performed in Kenya from 2012 to 2018. The previous evaluation points of the NSBDP were: baseline survey in 2012 (Year 1) [7], follow-up impact assessment one survey in 2016 (Year 3) [10], follow-up impact assessment two survey in 2017 (Year 5) [11], and follow-up impact assessment three survey in 2018 (Year 6) [4]. After ten years of the NSBDP's operation, there is a need to have a more granular understanding of the variation in helminth prevalence across Kenya. This includes, if and where, treatment may be suspended with appropriate surveillance implemented, as recommended by WHO, and how scarce resources can be best targeted to maximize the impact of the programme. To this end, the year nine (2021/2022) impact assessment four survey was conducted. Differing from previous survey designs, it made use of the previous surveys, environmental, treatment coverage, and spatial data to optimize the survey design and analysis.

1.4. Necessity for updated sampling methods and cost-effectiveness

PC is a cost-effective approach to controlling STH morbidity, but relies on large-scale surveys to determine and revise treatment frequency. Surveys represent a substantial proportion of helminth control programme budgets. As helminth prevalence reduces globally, there is a need to optimize cost components of these surveys to make the best use of available resources. Recent innovations in survey design using geospatial statistical methods, such as model-based geostatistics (MBG), to select survey sites have shown to deliver more precise results, given the same resources, than traditional design approaches [12].

Critically, for the cost-effectiveness of helminth control programmes, MBG differs from traditional survey design in the selection of sites for surveying and the information which is derived from those sites post-survey. Traditional design suggests randomization of sites for each survey across representative areas, such as ecological zones. MBG uses predictive models based on the results of earlier surveys to identify sites which provide the most predictive power for post-survey modelling of prevalence [12]. In this way, spatial sampling can be used to target the most informative sites and maximize survey precision under given resource constraints. MBG, uses post-survey information to create predictive models of prevalence that are as accurate as possible, under some stated assumptions. Such targeting also maximizes the potential of integration of STH and schistosomiasis, given their similarities in risk-factors. Finally, MBG allows the estimation of probabilities that post-analysis prevalence lie within pre-defined programmatically relevant thresholds [13].

1.5. Study objectives

The primary objective of the survey was to determine the probability of various levels of implementation units (IUs) lying within pre-defined STH prevalence ranges, namely: 0 % to <2 %, 2 % to <10 %, 10 % to <20 %, 20 % to <50 % and $\geq$ 50 %. The different IU levels included the overall NSBDP geographic area, county, sub-county, and ward. The secondary objective was to estimate IU level mean prevalence and intensity of infection with associated level of uncertainty.

2. Methods

2.1. Study design and sampling using MBG approach

The year 9 survey utilized a cross-sectional study design where 200 schools drawn from parts of seven regions of Kenya were sampled using a spatially regulated design [13]. This class of designs uses a constrained randomization that imposes a minimum distance between any two sampled locations (schools). In this survey, a minimum distance of 10 Km between schools was pre-specified. This constrained randomization results in a more even coverage of the geographical region of interest than would be obtained by an unconstrained randomization and usually leads to better predictive performance [14].

All the 27 NSBDP counties that are currently receiving treatment for STH were included in the survey as IUs. Spatially regulated sampling was conducted within each IU that selected schools. The number of schools selected per IU varied based on the risk profile of the IU as determined using the MBG approach. An IU is a geographical area over which a particular treatment strategy is applied (e.g., ward, sub-county, county, province or country). It is expected that after five to six years of consistent mass drug administration (MDA) the infection prevalence will substantially reduce and the STH prevalence classification will need to be reviewed with the aim of changing the treatment delivery frequency according to the WHO decision tree [15].

In each school, a minimum random sample of 70 children was taken. The sample included equal number of participants of each gender per class for seven classes, one early childhood development (ECD) class and classes one to six. Sample size calculations used the MBG approach to achieve a pre-specified proportion of 87 % of correctly classified IUs [13].

2.2. Survey procedures

The selected schools were visited three days prior to the survey date to brief the school head teacher and committee on the purpose of the survey. On the day of the survey, each selected child was given a container (poly-pot) labelled with a unique identifier and was instructed to place a portion of his or her own stool sample in it. The stool samples were then processed in the laboratory within 24 hours and examined in duplicate for the presence of STH eggs by two technicians using the Kato-Katz technique [16]. The survey procedures used during this survey, including the diagnosis technique, were similar to those used during previous impact assessments, and only the design and analysis of the survey that differed.

2.3. Data collection and management

The survey data were collected in two phases, both prior to the year 10 MDA. The phase one survey was conducted between 6th and 24th September, 2021, and phase two between 9th May and 22nd June, 2022. Data on the infection prevalence and intensity were collected by examining a single stool sample from each surveyed child and recording the number of STH eggs. The survey laboratory reporting form was programmed on to android-based smart phones which were used to capture data electronically using the Open Data Kit system that incorporated in-built data quality checks to reduce data entry errors [17]. Data was sent to a secure server in Nairobi, and processed as per the Kenya data protection legislation and institutional data policies and guidelines.

2.4. Statistical analysis and modelling

2.4.1. The geostatistical model

The data obtained from the i^{th} surveyed school are: n_i, the number of children tested; x_i, the school's location; and y_i, the number of children whose test result is positive, where "*positive*" can mean the detection of any eggs in the stool sample or of a number of eggs corresponding to light, moderate or heavy infection. Prevalence, defined as the probability of a positive test result, at any location x in the geographical region of interest A, is denoted by $P(x)$. Given $P(x_i)$, the probability distribution of y_i is binomial, with denominator n_i.

Finally, $P(x_i)$ is modelled as a mixed effects logistic regression,

$$log\left[\frac{P(x_i)}{1-P(x_i)}\right] = d(x_i)'\beta + S(x_i) + Z_i$$

where; $d(x)$ is a set of context-specific covariates associated with location x, β is the corresponding set of regression parameters, $S(x)$ is a spatially correlated Gaussian stochastic process and the Z_i are zero-mean Normally distributed random variables realised independently at each sampled location. The terms $d(x_i)'\beta$ and $S(x_i)$ represent spatial variation in prevalence that can and cannot, respectively, be explained by covariates that are available throughout the geographical region of interest, whilst the Z_i represent spatially unstructured extra-binomial variation at each sampled location, for example due to familial clustering of infection status [18]. A set of covariates $d(x_i)$ were included in the models used to design the year nine impact survey but not to analyse the collected data, as the very low prevalence did not allow reliably estimation of the association between outcome and explanatory variables.

2.4.2. The geostatistical prediction

The inferential target in each IU is the population-weighted prevalence,

Table 1
Number of schools, children examined by county, school level prevalence range (min-max), and county level prevalence estimates among school children in Kenya.

County	No. schools (No. children)	Median age (min-max)	School level STH prevalence[a] range (min-max)				County level prevalence estimate[a]			
			STH combined	Hookworm	*A. lumbricoides*	*T. trichiura*	STH combined	Hookworm	*A. lumbricoides*	*T. trichiura*
Bomet	7 (489)	9 (6–15)	0.0–18.8	0.0–2.9	0.0–15.9	0.0–1.4	8.3 (8.2–8.5)	0.5 (0.5–0.5)	7.3 (7.1–7.4)	0.7 (0.6–0.7)
Bungoma	9 (629)	10 (4–15)	1.4–50.7	0.0–1.5	1.4–34.8	0.0–15.9	16.9 (16.7–17.0)	0.3 (0.3–0.3)	14.2 (14.0–14.3)	2.9 (2.8–3.0)
Busia	7 (446)	9 (3–15)	0.0–10.1	0.0–1.4	0.0–4.3	0.0–7.2	6.3 (6.2–6.4)	0.5 (0.5–0.5)	4.3 (4.2–4.3)	1.6 (1.6–1.7)
Garissa	4 (237)	11 (5–16)	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0	2.3 (2.2–2.4)	0.2 (0.2–0.2)	0.9 (0.9–1.0)	1.2 (1.1–1.2)
Homabay	9 (618)	10 (4–16)	0.0–8.3	0.0–1.4	0.0–2.9	0.0–6.7	2.1 (2.1–2.2)	0.2 (0.2–0.2)	1.3 (1.2–1.3)	0.7 (0.6–0.7)
Kakamega	8 (523)	10 (5–14)	1.5–19.7	0.0–2.8	1.5–18.3	0.0–4.2	14.2 (14.0–14.4)	0.4 (0.4–0.4)	11.8 (11.6–11.9)	2.3 (2.3–2.4)
Kericho	5 (350)	9 (6–14)	0.0–10.0	0.0–0.0	0.0–10.0	0.0–1.4	5.0 (4.9–5.1)	0.2 (0.2–0.3)	4.0 (3.9–4.1)	0.8 (0.7–0.8)
Kilifi	8 (554)	10 (5–17)	0.0–2.9	0.0–1.4	0.0–1.4	0.0–0.0	0.9 (0.9–0.9)	0.2 (0.2–0.2)	0.4 (0.4–0.4)	0.3 (0.3–0.3)
Kirinyaga	2 (138)	8 (5–14)	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0	0.7 (0.7–0.8)	0.2 (0.2–0.2)	0.3 (0.3–0.3)	0.2 (0.2–0.3)
Kisii	4 (244)	9 (5–16)	1.4–12.9	0.0–2.9	1.4–11.4	0.0–2.9	12.3 (12.1–12.6)	0.2 (0.2–0.2)	10.1 (9.9–10.4)	2.3 (2.2–2.4)
Kisumu	8 (552)	9 (1–14)	0.0–21.7	0.0–0.0	0.0–14.5	0.0–13.0	4.9 (4.8–5.0)	0.2 (0.2–0.2)	3.1 (3.0–3.2)	1.7 (1.6–1.8)
Kitui	22 (1473)	9 (3–15)	0.0–2.0	0.0–0.0	0.0–1.9	0.0–2.0	0.9 (0.9–0.9)	0.2 (0.2–0.2)	0.4 (0.3–0.4)	0.4 (0.4–0.4)
Kwale	6 (372)	10 (4–19)	0.0–1.5	0.0–1.5	0.0–0.0	0.0–0.0	0.8 (0.8–0.9)	0.3 (0.3–0.3)	0.3 (0.3–0.3)	0.2 (0.2–0.3)
Lamu	3 (203)	9 (5–14)	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0	1.3 (1.3–1.4)	0.2 (0.2–0.2)	0.5 (0.5–0.6)	0.6 (0.5–0.7)
Machakos	15 (958)	8 (2–17)	0.0–11.4	0.0–2.9	0.0–8.6	0.0–2.9	1.0 (0.9–1.0)	0.2 (0.2–0.2)	0.6 (0.6–0.6)	0.2 (0.2–0.2)
Makueni	14 (936)	8 (4-`15)	0.0–5.7	0.0–1.5	0.0–5.7	0.0–2.9	1.6 (1.6–1.6)	0.2 (0.2–0.2)	1.1 (1.1–1.1)	0.3 (0.3–0.4)
Migori	8 (560)	9 (4–14)	0.0–10	0.0–1.4	0.0–10.0	0.0–0.0	4.9 (4.8–5.0)	0.2 (0.2–0.2)	4.3 (4.2–4.4)	0.5 (0.4–0.5)
Mombasa	2 (133)	10 (5–15)	0.0–1.4	0.0–0.0	0.0–1.4	0.0–0.0	0.7 (0.7–0.8)	0.2 (0.2–0.2)	0.4 (0.4–0.4)	0.1 (0.1–0.2)
Nandi	6 (411)	9 (2–16)	0.0–14.7	0.0–5.8	0.0–14.7	0.0–1.4	8.6 (8.5–8.8)	0.9 (0.8–0.9)	6.9 (6.8–7.1)	0.9 (0.9–1.0)
Narok	17 (1175)	10 (4–18)	0.0–71.4	0.0–3.0	0.0–24.3	0.0–67.1	10.8 (10.7–10.9)	0.2 (0.2–0.3)	6.2 (6.1–6.3)	5.0 (4.9–5.1)
Nyamira	3 (210)	9 (5–14)	8.6–21.4	0.0–0.0	8.6–18.6	0.0–2.9	15.8 (15.5–16.1)	0.2 (0.2–0.2)	14.1 (13.7–14.4)	1.9 (1.7–2.0)
Siaya	7 (483)	9 (4–15)	0.0–17.1	0.0–2.9	0.0–10.0	0.0–11.8	10.7 (10.5–10.8)	0.5 (0.5–0.5)	5.0 (4.9–5.0)	5.6 (5.5–5.8)
Taita Taveta	5 (344)	9 (5–15)	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0	0.7 (0.6–0.7)	0.2 (0.2–0.2)	0.2 (0.2–0.2)	0.3 (0.2–0.3)
Tana River	4 (250)	11 (3–18)	0.0–8.7	0.0–0.0	0.0–4.3	0.0–5.1	3.9 (3.8–4.0)	0.2 (0.2–0.2)	1.6 (1.5–1.6)	2.2 (2.1–2.3)
Trans Nzoia	8 (552)	10 (1–15)	2.9–12.9	0.0–1.4	2.9–11.4	0.0–2.9	9.1 (9.0–9.2)	0.2 (0.2–0.2)	7.6 (7.5–7.7)	1.4 (1.4–1.5)
Vihiga	1 (70)	10 (4–15)	7.1–7.1	0.0–0.0	5.7–5.7	1.4–1.4	10.8 (10.5–11.1)	0.3 (0.3–0.3)	7.4 (7.1–7.6)	3.5 (3.3–3.7)
Wajir	8 (506)	11 (4–27)	0.0–1.7	0.0–1.7	0.0–1.5	0.0–0.0	1.9 (1.8–2.0)	0.3 (0.3–0.3)	0.8 (0.8–0.9)	0.8 (0.8–0.9)
Total	**200 (13,416)**	**9 (1–27)**	**0.0–71.4**	**0.0–5.8**	**0.0–34.8**	**0.0–67.1**	**5.8 (5.7–6.0)**	**0.3 (0.2–0.4)**	**4.3 (4.2–4.4)**	**1.4 (1.3–1.5)**

[a] Prevalence was calculated using a model-based geostatistical approach that accounted for both the observed explanatory variables and the unobserved stochastic processes around a specific location.

$$T = \int pd(x)P(x)dx$$

where $pd(x)$ is the population density at x and the integration is over the whole of the IU in question. If the locations and enrolments of all schools in the IU are known, the integral would reduce to a sum; here, for $pd(x)$ we use population density data on a regular grid at a spacing of 5 Km, which we obtained from WorldPop [19]. After the model parameters have been estimated by Monte Carlo maximum likelihood, we draw samples from the predictive distribution of T, from which we calculate the probability distribution of T over the designated set of prevalence intervals and classify the IU accordingly. Technical details are given in the appendix A.

3. Results

During year nine survey, 200 schools and 13,416 children with median age of 9 years (range: 1–19 years) were surveyed across all the 27 NSBDP counties. Approximately half 6790 (50.6 %) of the surveyed children were males. Distribution of the children per class was as follows: ECD 1863 (13.9 %), class one 842 (6.2 %), class two 2635 (19.6 %), class three 2636 (19.7 %), class four 2646 (19.7 %), class five 2625 (19.6 %), and class six 169 (1.3 %). The number of schools and children surveyed varied per county are indicated in Table 1.

3.1. Predictive probability estimates of STH prevalence intervals using MBG approach

The predictive probabilities of any STH prevalence lying within 0 % to <2 %, 2 % to <10 %, 10 % to <20 %, 20 % to <50 % and ≥ 50 % were determined. Probabilities were estimated at the level of the NSBDP geographic area, counties and sub-counties. A given geographic unit was assigned to a particular threshold if its predictive probability was greater than 0.500. All thresholds assigned at county-level are in Table 2, while an overview of thresholds assigned to sub-counties is shown in Fig. 1.

The overall NSBDP geographic area prevalence for any STH infection was estimated to sit within 2 % to <10 % with probability >0.999. Species-specific classifications were 2 % to <10 % for *A. lumbricoides,* 0 % to <2 % for hookworm and 0 % to <2 % for T. *trichiura* all with probability >0.999. County-specific thresholds showed some variation. Of the 27 NSBDP counties, ten counties were assigned to sit within 0 % to <2 % prevalence (with >0.999 predictive probability except for Wajir with predictive probability 0.971). Another ten counties were assigned to sit within 2 % to <10 % prevalence with probability >0.999, and seven counties were assigned

Table 2
STH county endemicity classification using predictive probabilities and the number of sub-counties classified according to their respective county STH endemicity calculated from the fitted MBG model.

County	Mean prevalence estimate (%)	Predictive probability of classifying a county to a given STH endemicity class					Total sub-counties estimated	Total number of sub-counties in each county classified according to STH endemicity class				
		<2 %	2–10 %	10–20 %	20–50 %	>50 %		<2 %	2–10 %	10–20 %	20–50 %	>50 %
Bomet	8.317	<0.001	**>0.999**	<0.001	<0.001	<0.001	5	0	4	1	0	0
Bungoma	16.856	<0.001	<0.001	**>0.999**	<0.001	<0.001	8	0	2	2	4	0
Busia	6.318	<0.001	**>0.999**	<0.001	<0.001	<0.001	7	0	7	0	0	0
Garissa	2.322	<0.001	**>0.999**	<0.001	<0.001	<0.001	6	3	3	0	0	0
Homa Bay	2.136	<0.001	**>0.999**	<0.001	<0.001	<0.001	8	5	3	0	0	0
Kakamega	14.192	<0.001	<0.001	**>0.999**	<0.001	<0.001	12	0	2	9	1	0
Kericho	4.983	<0.001	**>0.999**	<0.001	<0.001	<0.001	6	0	6	0	0	0
Kilifi	0.898	**>0.999**	<0.001	<0.001	<0.001	<0.001	7	7	0	0	0	0
Kirinyaga	0.736	**>0.999**	<0.001	<0.001	<0.001	<0.001	4	4	0	0	0	0
Kisii	12.311	<0.001	<0.001	**>0.999**	<0.001	<0.001	9	0	5	3	1	0
Kisumu	4.929	<0.001	**>0.999**	<0.001	<0.001	<0.001	8	0	8	0	0	0
Kitui	0.902	**>0.999**	<0.001	<0.001	<0.001	<0.001	8	8	0	0	0	0
Kwale	0.839	**>0.999**	<0.001	<0.001	<0.001	<0.001	4	4	0	0	0	0
Lamu	1.322	**>0.999**	<0.001	<0.001	<0.001	<0.001	2	1	1	0	0	0
Machakos	0.958	**>0.999**	<0.001	<0.001	<0.001	<0.001	8	8	0	0	0	0
Makueni	1.601	**>0.999**	<0.001	<0.001	<0.001	<0.001	6	5	1	0	0	0
Migori	4.891	<0.001	**>0.999**	<0.001	<0.001	<0.001	8	1	7	0	0	0
Mombasa	0.747	**>0.999**	<0.001	<0.001	<0.001	<0.001	6	6	0	0	0	0
Nandi	8.616	<0.001	**>0.999**	<0.001	<0.001	<0.001	6	0	5	1	0	0
Narok	10.815	<0.001	<0.001	**>0.999**	<0.001	<0.001	6	1	3	1	1	0
Nyamira	15.810	<0.001	<0.001	**>0.999**	<0.001	<0.001	4	0	0	4	0	0
Siaya	10.679	<0.001	<0.001	**>0.999**	<0.001	<0.001	6	0	2	4	0	0
Taita Taveta	0.683	**>0.999**	<0.001	<0.001	<0.001	<0.001	4	4	0	0	0	0
Tana River	3.915	<0.001	**>0.999**	<0.001	<0.001	<0.001	3	1	2	0	0	0
Trans Nzoia	9.101	<0.001	**>0.999**	<0.001	<0.001	<0.001	5	0	3	2	0	0
Vihiga	10.825	<0.001	<0.001	**>0.999**	<0.001	<0.001	5	0	2	3	0	0
Wajir	1.925	**0.971**	0.029	<0.001	<0.001	<0.001	6	4	2	0	0	0

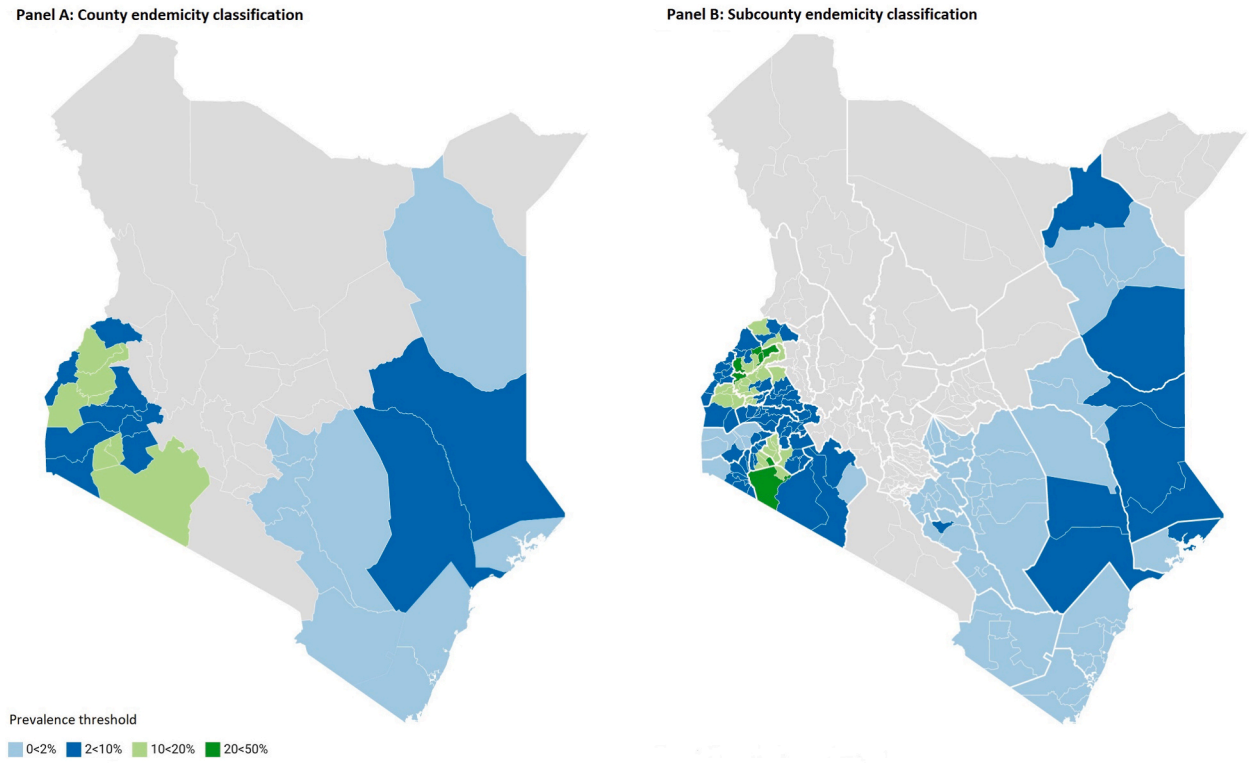

Fig. 1. STH county (panel A) and sub-county (panel B) endemicity classification using predictive probabilities calculated from the fitted MBG model.

to sit within 10 % to <20 % with probability >0.999.

As described above, the analysis was extended to determine sub-county variation in prevalence thresholds. It should be noted that with more geographic focality, predictions are necessarily less precise. The analysis suggested that in 17 of the 27 NSBDP counties, the assigned prevalence categories varied across sub-counties. For example, of the nine sub-counties which comprise Bungoma County, three were assigned to sit within 2 % to <10 %, two within 10 % to <20 %, and four within 20 % to <50 %. Table 2 shows the number of sub-counties classified according to their county endemicity lying within specified thresholds.

3.2. Estimation of STH infection prevalence and uncertainty intervals using MBG approach

Using MBG approach, the overall STH prevalence was estimated to be 5.8 % (95%CI: 5.7–6.0) with species-specific prevalence of 4.3 % (95%CI: 4.2–4.4) for *A. lumbricoides*, 0.3 % (95%CI: 0.2–0.4) for hookworm, and 1.4 % (95%CI: 1.3–1.5) for *T. trichiura* (Table 1). County level prevalence ranged from 0.7 % to 16.9 % for any STH, 0.2 %–14.2 % for *A. lumbricoides*, 0.1 %–5.6 % for *T. trichiura*, and 0.2 %–0.5 % for hookworm (Table 1). The pixel level geographical distribution of STH prevalence across the NSBDP geographic area is shown in Fig. 2.

3.3. Estimation of the prevalence of moderate-to-heavy STH infection using classical statistical approach

The criteria for EPHP is that the prevalence of moderate-to-heavy intensity of STH infection is less than 2 %. The very low frequency of moderate-to-heavy intensity of infection in the data prevents the use of the standard geostatistical modelling approach. Therefore, we used only a classical statistical approach by assuming a binomial sampling distribution for the number of moderate-to-heavy infections with denominator taken as the total number of children examined.

The overall prevalence of moderate-to-heavy intensity was 1.3 % (95%CI: 1.1–1.5). The comparison of the prevalence of moderate-to-heavy intensity of STH infection between baseline and year nine evaluation is given in Fig. 3 panel A.

The county-level prevalence of moderate-to-heavy intensity of any STH infection is shown in Fig. 3 panel B; which varied from 0 % to 8.1 %. From these results, 8 out of the 27 counties have not eliminated STH as a public health problem.

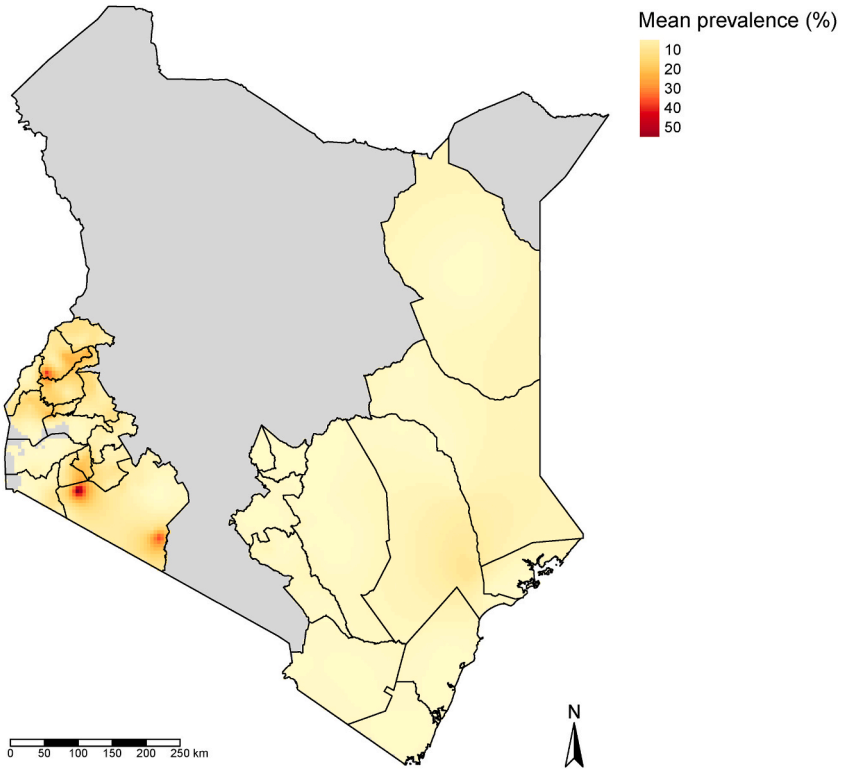

Fig. 2. Pixel level geographical distribution of STH mean prevalence estimated from the fitted MBG model.

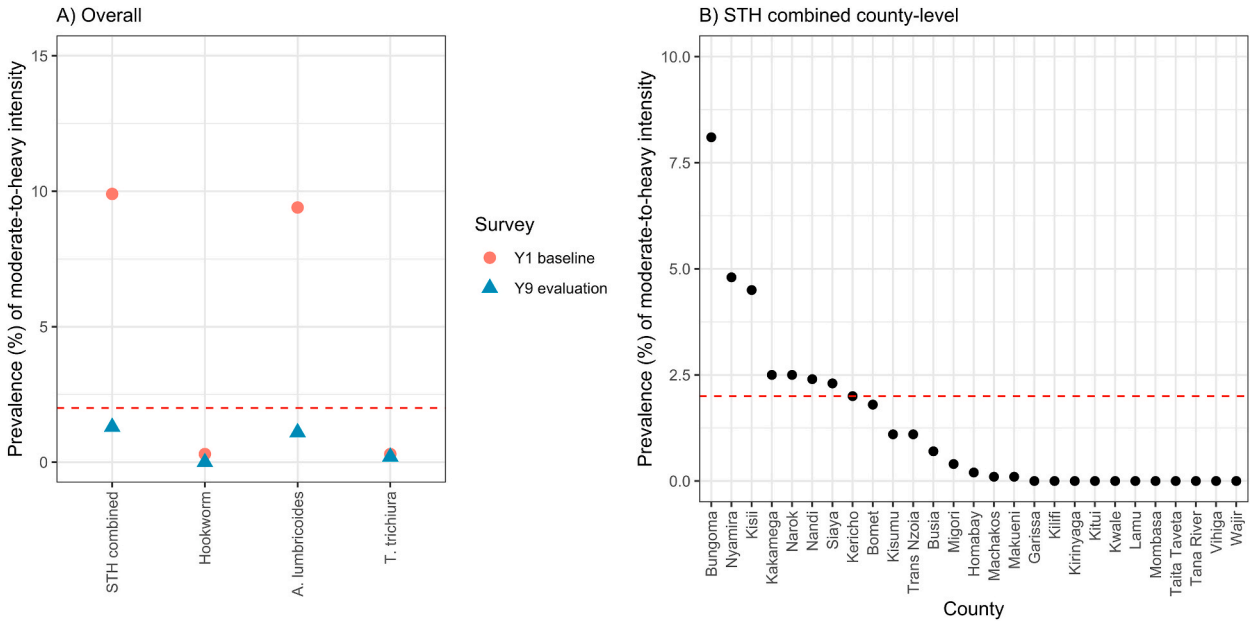

Fig. 3. The overall (panel A) and county level (panel B) prevalence of moderate-to-heavy intensity of STH infection calculated using classical statistical models with the total number of children examined taken as a denominator. The red dotted lines indicate the cut-off level (2.0 %) for the prevalence of moderate-to-heavy intensity below which the NSBDP geographic area or a county is considered to have attained EPHP. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

The current WHO guidelines states that the STH prevalence in a given geographical area should be estimated based on population surveys [20]. This is then used to determine whether MDA should be initiated, and if so, at what frequency. Additionally, the same guidelines state that for areas where MDA has been delivered consistently for about five to six years, it is necessary to conduct an impact assessment in order to further refine MDA requirements. In this study, we used the novel MBG approach to design and analyse a large-scale nation-wide impact assessment survey with the aim of accurately determining the probability that the overall NSBDP geographic area, counties and sub-counties lie within programmatically relevant prevalence thresholds. In this way, we are able to confidently refine the MDA requirements for the programme areas.

Prevalence mapping and impact assessment typically rely on a collection of empirical data using finite, but often spatially sparse, set of surveys of communities within the region of interest [21]. The design of these surveys often follows WHO recommendations to sample 50 children per school with the number of schools to be surveyed, limited to the available budget, and selection of school locations following random sampling designs [22]. The analysis of the prevalence data from these surveys are often performed using standard models of the form of a generalized linear mixed model with binomial error distribution, logistic link and a set of explanatory variables [23]. The estimates resulting from the standard models are mostly unbiased and inefficient, provided that the participants (both schools and children) were randomly selected from the general population [22].

For the first time in a national survey, predictive probabilities are used to allocate treatment requirements while taking into account the varying geography. In this survey, the results indicated that the overall NSBDP geographic area prevalence for any STH sit between 2 % and <10 % with a sufficiently high predictive probability of >0.999. Typically, the whole of this geographic area would need to be treated once every two years, per the WHO guidelines [24]. However, considering the heterogeneity in infection burden within counties, the analysis was powered to provide predictive probabilities at county level. As such, out of the 27 counties, 10 counties which include Kilifi, Kirinyaga, Kitui, Kwale, Lamu, Machakos, Makueni, Mombasa, Taita Taveta, and Wajir were estimated with high predictive probability to sit within 0 % to <2 % prevalence. This is strong evidence that the STH prevalence in these counties is below the MDA threshold set by WHO and therefore these counties should consider suspending treatment and implement appropriate surveillance system. Another ten counties which include Bomet, Busia, Garissa, Homa Bay, Kericho, Kisumu, Migori, Nandi, Tana River, and Trans Nzoia had their prevalence estimated with a high predictive probability to sit within 2 % to <10 %, and as such require MDA only once after every two years. Lastly, seven counties which include Bungoma, Kakamega, Kisii, Narok, Nyamira, Siaya and Vihiga had their prevalence estimated with a high predictive probability to sit within 10 % to <20 %. Therefore, these counties should consider maintaining the previous treatment plan of annual MDA. Further, even though the analysis was extended to determine subcounty heterogeneity in prevalence thresholds, treatment decisions were made at county level while taking into account sub-counties with high predictive probabilities. This was because the more geographic focality, the less confidence can be placed on the assumptions made to derive prevalence thresholds and predictive probabilities at subcounty levels [12].

After nine rounds of consistent MDA in the 27 NSBDP counties, the overall prevalence of any STH reduced to low levels of 5.8 % down from initial level of 32.3 % [7]. However, the prevalence is still above the no MDA requirement threshold of <2 % [24]. Additionally, the county and subcounty levels prevalence is heterogeneous with some counties showing up to 16.9 % prevalence. Hence, the need to continue with MDA but with a county-focused frequency strategy. This kind of treatment strategy changes will help ease the scarce resources and target the reservoirs of continued disease transmission in order to expedite programme impact. Counties where treatment may be suspended are recommended to implement surveillance systems as per the WHO guidelines.

The prevalence of moderate to heavy intensity of infection is an important metric in STH control since it indicates whether the programme has achieved EPHP, which is defined as a geographic area having reached less than 2 % of moderate to heavy intensity of STH infection [25]. In view of the NSBDP geographic area (consisting of 27 counties), EPHP should be evaluated within each specific county. Whilst, the overall prevalence of moderate to heavy intensity was 1.3 %, which is below the threshold indicating attainment of EPHP, the county level EPHP showed variation. It is important to highlight that EPHP should be declared only if all the IUs (in this case counties) have reached the below threshold. Out of the 27 counties surveyed, 19 counties have attained EPHP while 8 counties which include Bungoma (8.1 %), Nyamira (4.8 %), Kisii (4.5 %), Narok (2.5 %), Kakamega (2.5 %), Nandi (2.4 %), Siaya (2.3 %) and Kericho (2.0 %), have not attained the threshold and shows relatively high disease burden. Previous surveys have documented challenges in the same counties related to transmission drivers including poor sanitation and hygiene conditions, slightly low treatment coverage, suitable environmental conditions for parasite persistence and economic activities that expose individuals to parasite interactions, among other factors [26]. Improved interventions targeting attainment of EPHP in these counties should be explored as well as research studies to document reasons of persistence of STH in these areas.

4.1. Study strengths and cost-effectiveness of the MBG approach

The MBG derive its cost-effectiveness from the fact that; (i) pre-intervention prevalence survey information can be used to design stratified impact surveys in which areas of historically low prevalence can be under-sampled, resulting in a more efficient use of scarce resources [22], (ii) the design balance between best possible predictive performance at an affordable cost (cost-effectiveness) and an acceptable predictive performance (precision), and (iii) the design of MBG can be optimized to target more than one NTD or other diseases provided they inherit similar spatial correlation properties, giving effect to integrated survey design approach that then make prudent use of the limited resources typical of settings in low- and middle-income countries [27]. However, the main limitation of the MBG is that for effective widespread application, it requires additional statistical analysis skills.

5. Conclusions

The findings from this impact assessment survey showed a continued reduction in STH prevalence since the start of the NSBDP in 2012. Based on these results, morbidity due to STH is no longer an issue at the population level among the general SAC population, this however, may not be the case when monitoring the infection at specific county level. Using the MBG approach, the STH treatment requirements can now be confidently refined. For STH, the prior approach has been to target all the NSBDP geographic areas for annual treatment. However, the results showed that, currently, a number of counties have a very high predictive probability of sitting below 2 % or within 2 % to 10 % prevalence. As such the programme should adopt county-level treatment frequencies and consider suspension of treatment in counties that have reduced prevalence to below threshold (i.e., <2 % prevalence) while establishing a cost-effective surveillance system in those counties.

Funding source

The financial support for this study was provided by The END Fund, GiveWell and Epic Foundation via Evidence Action. Collins Okoyo is supported by Evidence Action through Kenya Medical Research Institute as a Principal Statistician, and further supported by GlaxoSmithKline (GSK) Africa Non-Communicable Disease Open Lab through the DELTAS Africa Sub-Saharan African Consortium for Advanced Biostatistics (SSACAB) training programme, grant number D1702270-01, as a PhD fellow. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Ethical approval statement

Ethical approval for the study protocol was obtained from the KEMRI Scientific and Ethics Review Unit (SERU Number 2206). Permission to access the schools for the survey was obtained from the national-level Ministry of Education. At county-level, approval was provided by the respective county health and education authorities. At school-level, parental consent from parents/guardians of the children was based on passive, opt-out consent rather than written opt-in consent due to the low-risk nature of the survey procedure. Additionally, individual assent was obtained from each child before participation in the study. All data used were anonymized.

Data sharing

The de-identified school level dataset used in this analysis can be obtained from the supplementary file in the Appendix A. For individual level de-identified data, reasonable request can be made via email to the corresponding author.

Data availability statement

The analysed data is provided with this submission as a supplementary material, or can be accessed at https://github.com/mancollo/NSBDP_2022_School_Level_Dataset. Date October 2, 2023.

CRediT authorship contribution statement

Collins Okoyo: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draftWriting – original draft, Writing – review & editingWriting – review & editing. **Mark Minnery:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – original draftWriting – original draft, Writing – review & editingWriting – review & editing. **Idah Orowe:** Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Chrispin Owaga:** Conceptualization, Data curation, Investigation, Project administration, Supervision, Validation, Writing – review & editingWriting – review & editing. **Suzy J. Campbell:** Investigation, Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Christin Wambugu:** Data curation, Investigation, Supervision, Validation, Writing – review & editingWriting – review & editing. **Nereah Olick:** Data curation, Investigation, Supervision, Validation, Writing – review & editingWriting – review & editing. **Jane Hagemann:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Wyckliff P. Omondi:** Data curation, Investigation, Supervision, Validation, Writing – review & editingWriting – review & editing. **Kate McCracken:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Antonio Montresor:** Investigation, Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Graham F. Medley:** Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editingWriting – review & editing. **Claudio Fronterre:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – review & editingWriting – review & editing. **Peter Diggle:** Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – review & editingWriting – review & editing. **Charles Mwandawiro:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editingWriting – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to the National Ministry of Health and Ministry of Education, Kenya, and the respective County Ministries of Health and Education for their support, facilitation and guidance during implementation of this work. Additional thank you to the sub-county health management teams, sub-county education officials, teachers, children and their parents/guardians for their tireless effort and cooperation throughout the study. We are deeply indebted to all the study team members, field personnel and the entire staff at the Eastern and Southern Africa Centre of International Parasite Control (ESACIPAC), Kenya Medical Research Institute (KEMRI) for your efforts and dedication towards the success of this programme.

Appendix A. Supplementary data and technical details to statistical modeling

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e20695.

References

[1] N.R. de Silva, S. Brooker, P.J. Hotez, A. Montresor, D. Engels, L. Savioli, Soil-transmitted helminth infections: Updating the global picture, Elsevier Ltd, Trends Parasitol. 19 (12) (2003) 547–551, https://doi.org/10.1016/j.pt.2003.10.002. Dec. 01.
[2] J.E. Weatherhead, P.J. Hotez, R. Mejia, The global state of helminth control and elimination in children, Pediatr. Clin. 64 (4) (2017) 867, https://doi.org/10.1016/J.PCL.2017.03.005. Aug.
[3] S. Brooker, A.C.A. Clements, D.A.P. Bundy, Global epidemiology, ecology and control of soil-transmitted helminth infections, Adv. Parasitol. 62 (2006) 221–261, https://doi.org/10.1016/S0065-308X(05)62007-6.
[4] C. Okoyo, S.J. Campbell, K. Williams, E. Simiyu, C. Owaga, C. Mwandawiro, Prevalence, intensity and associated risk factors of soil-transmitted helminth and schistosome infections in Kenya: impact assessment after five rounds of mass drug administration in Kenya, PLoS Negl Trop Dis 14 (10) (Oct. 2020) 1–33, https://doi.org/10.1371/journal.pntd.0008604.
[5] W. Moser, C. Schindler, J. Keiser, Efficacy of recommended drugs against soil transmitted helminths: systematic review and network meta-analysis, The BMJ 358 (2017), https://doi.org/10.1136/BMJ.J4307.
[6] C. Urbani, M. Albonico, Anthelminthic drug safety and drug administration in the control of soil-transmitted helminthiasis in community campaigns, Acta Trop. 86 (2–3) (May 2003) 215–221, https://doi.org/10.1016/S0001-706X(03)00036-6.
[7] C.S. Mwandawiro, et al., Monitoring and evaluating the impact of national school-based deworming in Kenya: study design and baseline results, Parasit Vectors 6 (1) (2013) 198, https://doi.org/10.1186/1756-3305-6-198.
[8] R.M. Anderson, J.E. Truscott, R.L. Pullan, S.J. Brooker, T.D. Hollingsworth, How effective is school-based deworming for the community-wide control of soil-transmitted helminths? PLoS Negl Trop Dis 7 (2) (2013) https://doi.org/10.1371/JOURNAL.PNTD.0002027.
[9] M. Bangert, D.H. Molyneux, S.W. Lindsay, C. Fitzpatrick, D. Engels, The cross-cutting contribution of the end of neglected tropical diseases to the sustainable development goals, Infectious Diseases of Poverty 6 (2017), https://doi.org/10.1186/s40249-017-0288-0.
[10] C. Okoyo, et al., "Monitoring the impact of a national school based deworming programme on soil-transmitted helminths in Kenya: the first three years, 2012 - 2014,", Parasit Vectors 9 (1) (2016) https://doi.org/10.1186/s13071-016-1679-y.
[11] C. Mwandawiro, et al., Results of a national school-based deworming programme on soil-transmitted helminths infections and schistosomiasis in Kenya: 2012-2017, Parasit Vectors 12 (1) (2019), https://doi.org/10.1186/s13071-019-3322-1.
[12] P.J. Diggle, E. Giorgi, Model-based geostatistics for global public health: methods and applications, Model-based Geostatistics for Global Public Health: Methods and Applications (Mar. 2019) 1–247, https://doi.org/10.1201/9781315188492/MODEL-BASED-. GEOSTATISTICS-GLOBAL-PUBLIC-HEALTH-PETER-DIGGLE-EMANUELE-GIORGI.
[13] O. Johnson, et al., Model-based geostatistical methods enable efficient design and analysis of prevalence surveys for soil-transmitted helminth infection and other neglected tropical diseases, in: Clinical Infectious Diseases, vol. 72, Supplement_3. Oxford Academic, 2021, https://doi.org/10.1093/cid/ciab192. S172–S179, Jun. 14.
[14] M. Chipeta, D. Terlouw, K. Phiri, P.D. Environmetrics, undefined, Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure, Wiley Online Library 28 (1) (2017), https://doi.org/10.1002/env.2425. Feb. 2016.
[15] WHO, "2030 targets for soil-transmitted helminthiases control programmes,", Accessed: Mar. 11, 2020. [Online]. Available: http://apps.who.int/bookorders, 2020.
[16] J. Katz, N. Chaves, A. Pellegrino, A simple device for quantitative stool thick smear technique in Schistosomiasis mansoni, July, Rev. Soc. Bras. Med. Trop. 14 (1972) 397–400, 16167509.
[17] Open Odk, Data Kit, 2009. https://opendatakit.org/. (Accessed 18 February 2018).
[18] P.J. Diggle, E. Giorgi, Model-based Geostatistics for Global Public Health: Methods and Applications, CRC Press, Boca Raton, FL, USA, 2019, https://doi.org/10.1201/9781315188492.
[19] WorldPop, "Open spatial demographic data and research,", WorldPop, 2023. Apr. 29, https://www.worldpop.org. (Accessed 5 October 2023).
[20] WHO, Helminth Control in School-Age Children. A Guide for Managers of Control Programmes, World Health Organization, 2011. http://apps.who.int/iris/bitstream/10665/44671/1/9789241548267_eng.pdf. (Accessed 27 July 2017).
[21] P.J. Diggle, E. Giorgi, "Model-Based Geostatistics for Prevalence Mapping in Low-Resource Settings," 111 (515) (2016) 1096–1120, https://doi.org/10.1080/01621459.2015.1123158. Jul.
[22] O. Johnson, et al., Model-based geostatistical methods enable efficient design and analysis of prevalence surveys for soil-transmitted helminth infection and other neglected tropical diseases, Clin. Infect. Dis. 72 (Suppl 3) (Jun. 2021) S172–S179, https://doi.org/10.1093/CID/CIAB192.
[23] L. Lin, H. Chu, Meta-analysis of proportions using generalized linear mixed models, Epidemiology 31 (5) (Sep. 2020) 713–717, https://doi.org/10.1097/EDE.0000000000001232.
[24] WHO, Helminth Control in School Age Children: a Guide for Managers of Control Programmes, second ed., 2011. Geneva, Switzerland.

[25] A. Montresor, et al., The global progress of soil-transmitted helminthiases control in 2020 and World Health Organization targets for 2030, PLoS Negl Trop Dis 14 (8) (2020), e0008505, https://doi.org/10.1371/JOURNAL.PNTD.0008505. Aug.
[26] C. Okoyo, S.J. Campbell, C. Owaga, N. Onyango, G. Medley, C. Mwandawiro, Statistical regression model of water, sanitation, and hygiene; treatment coverage; and environmental influences on school-level soil-transmitted helminths and schistosome prevalence in Kenya: secondary analysis of the national deworming program data, Am. J. Trop. Med. Hyg. 104 (6) (Jun. 2021) 2251, https://doi.org/10.4269/AJTMH.20-1189.
[27] MG Chipeta, DJ Terlouw, KS Phiri, PJ Diggle, Adaptive geostatistical design and analysis for prevalence surveys, Spatial Statistics 15 (2016) 70–84.