RESEARCH ARTICLE

Appropriateness of the current parasitological control target for hookworm morbidity: A statistical analysis of individual-level data

Veronica Malizia[1,2]*, **Federica Giardina**[1,2], **Sake J. de Vlas**[1], **Luc E. Coffeng**[1]*

1 Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, **2** Department for Health Evidence, Biostatistics Research Group, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

* veronica.malizia@radboudumc.nl (VM); l.coffeng@erasmusmc.nl (LEC)

Abstract

Background

Soil-transmitted helminths affect almost 2 billion people globally. Hookworm species contribute to most of the related morbidity. Hookworms mainly cause anaemia, due to blood loss at the site of the attachment of the adult worms to the human intestinal mucosa. The World Health Organization (WHO) aims to eliminate hookworm morbidity by 2030 through achieving a prevalence of moderate and heavy intensity (M&HI) infections below 2%. In this paper, we aim to assess the suitability of this threshold to reflect hookworm-attributable morbidity.

Methodology/Principal findings

We developed a hierarchical statistical model to simulate individual haemoglobin concentrations in association with hookworm burdens, accounting for low haemoglobin values attributable to other causes. The model was fitted to individual-level data within a Bayesian framework. Then, we generated different endemicity settings corresponding to infection prevalence ranging from 10% to 90% (0% to 55% M&HI prevalence), using 1, 2 or 4 Kato-Katz slides. For each scenario, we estimated the prevalence of anaemia due to hookworm. Our results showed that on average, haemoglobin falls below the WHO threshold for anaemia when intensities are above 2000 eggs per gram of faeces. For the different simulated scenarios, the estimated prevalence of anaemia attributable to hookworm ranges from 0% to 30% (95%-PI: 24% - 36%) being mainly associated to the prevalence of M&HI infections. Simulations show that a 2% prevalence of M&HI infections in adults corresponds to a prevalence of hookworm-attributable anaemia lower than 1%.

Conclusions/Significance

Our results support the use of the current WHO thresholds of 2% prevalence of M&HI as a proxy for hookworm morbidity. A single Kato-Katz slide may be sufficient to assess the achievement of the morbidity target. Further studies are needed to elucidate haemoglobin dynamics pre- and post- control, ideally using longitudinal data in adults and children.

Author summary

Soil-transmitted helminths affect almost 2 billion people globally. About 65% of the related morbidity is attributable to hookworm infection, especially anaemia, which is caused by adult worms attaching to the intestinal mucosa of the human host to feed on blood. The World Health Organization (WHO) defines the target for the elimination of morbidity as reaching moderate-to-heavy intensity infections prevalence < 2%. In this paper, we aim to assess the suitability of the current WHO threshold to reflect hookworm-attributable anaemia. To this end, we developed a statistical model to simulate individual haemoglobin concentrations in association with hookworm burdens in adults, accounting for low haemoglobin values attributable to other causes, such as malnutrition and malaria. We used model outcomes to estimate the prevalence of hookworm-attributable anaemia in different endemicity scenarios. Our predictions suggested that anaemia is mainly associated to moderate-to-heavy intensity infection and that the 2% threshold corresponds to a prevalence of hookworm-attributable anaemia <1%. These results support the use of the current WHO threshold as a proxy for hookworm morbidity. Our predictions represent a first step towards better quantifying the prevalence of anaemia due to hookworm infection.

Introduction

Soil-transmitted helminths (STH) are estimated to affect almost 2 billion people globally [1], with the vast majority living in developing countries. Transmitted through contaminated soil, STH are responsible for considerable morbidity. About 65% of the overall estimated STH morbidity is attributable to hookworm infection, which is commonly caused by the helminth parasites *Necator americanus* and *Ancylostoma duodenale* [2]. The most common condition associated with hookworm infection is anaemia resulting from adult hookworms attaching to the mucosa of the small intestine in the human host to feed on blood. The uptake of blood and the extra loss due to wounds left by detaching worms reduce haemoglobin (Hb) and iron levels of the host, until the nutritional iron reserves and intake are not sufficient anymore [3]. Hookworm infection has been shown to significantly contribute to the global burden of anaemia in many countries [4].

The World Health Organization (WHO) categorises hookworm infections as light ($\leq$ 1999 eggs per gram of faeces (epg)), moderate (2000–3999 epg) and heavy ($\geq$ 4000 epg) [5]. The current global target set by the WHO is the elimination of hookworm as a public health problem by 2030, defined as reaching a moderate-to-heavy intensity (M&HI) prevalence < 2% in school-aged children [6]. Although anaemia is thought to be a consequence of M&HI infection, lower Hb levels in adults have been associated with all levels of infection intensity [7]. The number of adult worms in the human host that are sufficient to cause anaemia depends on several factors, such as the hookworm species causing the infection, the underlying nutritional reserves of the host, and coinfection with other parasites [8]. Clear evidence for the definition of the WHO guidelines and the use of M&HI prevalence of infection as indicator of hookworm morbidity is currently lacking.

Previous studies estimated the effect of hookworm infection intensity on blood iron reserves [3] and investigated the relationship between Hb and faecal egg counts [9]. For instance, Lwambo et al. [9] showed that Hb significantly declines for infection intensities above 2000 epg. They also predicted a non-linear relationship between prevalence of infection

and prevalence of hookworm-related anaemia where the latter steeply increases with high prevalence of infection. In these studies, results are presented in terms of overall prevalence of anaemia in the population. However, in most of the settings where hookworm infection is prevalent, malnutrition and coinfection with other endemic infectious diseases (e.g., malaria, schistosomiasis, other neglected tropical diseases) are also common. Therefore, it is important to account for an underlying prevalence of anaemia in the population due to causes other than hookworm infection.

The aim of this paper is to assess the suitability of the current WHO targets using prevalence of M&HI infection as a proxy for hookworm morbidity. To do this, we estimate the prevalence of anaemia attributable to hookworm infection as a function of hookworm prevalence (M&HI and any intensity). We do so by modelling how individual level Hb concentrations change in response to individual hookworm burden, indirectly observed through excreted egg counts. We account for anaemia attributable to other causes by assuming an Hb distribution that allows low values also in absence of worms. To extend this approach to other settings, we generate different endemicity scenarios and for each we estimate the prevalence of anaemia attributable to hookworm. We further consider the impact of different sampling schemes for diagnosis using the Kato-Katz (KK) thick smear technique on the evaluation of the morbidity target.

Methods

Ethics statement

The data employed in the present study were collected for previous publication [10]. The original study protocol was approved by the Makerere University Faculty of Medicine Research and Ethics Committee (#2008–043), the Uganda National Council of Science and Technology (#HS 476) and the London School of Hygiene and Tropical Medicine Ethics Committee (#5261) [10].

Data

We analysed individual data from a rural community in Uganda (N = 2037 individuals), consisting of Hb values and egg count measurements for hookworm and *Schistosoma mansoni* detection, collected by means of two KK slides per day per individual, for a total of two consecutive days and four slides for each parasite. Age, sex, and geographical location were also recorded. These data have been previously collected and analysed to estimate the spatial and genetic variation in intensity of hookworm infection [10]. Hb concentrations and egg counts for hookworm were available for 1840 individuals. The WHO provides the definition of anaemia based on age- and sex-specific thresholds [11]: women and men > 15 years old are defined anaemic if their Hb concentration is below 120 and 130 g/L, respectively. Due to high age/sex-variation in anaemia thresholds for children and the assumption that exposure to hookworm infection stabilises in adult age [12], we focused our analysis on the adult population ($\geq$ 20 years, both sexes). In ten individuals, extremely high egg counts were recorded (up to a maximum of 35,000 epg). We hypothesised that these few individuals were infected with the hookworm species *A. duodenale* and therefore excluded them (individuals with a mean egg count > 10,000 epg) from the analysis, due to higher range of daily egg production [8] and contribution to anaemia for *A. duodenale* [13]. The final population is composed of N = 695 adults with a 52% prevalence of any intensity of hookworm infection, 4% prevalence of M&HI of infection and 41% prevalence of overall anaemia. The process of data selection is summarised in **Fig 1**.
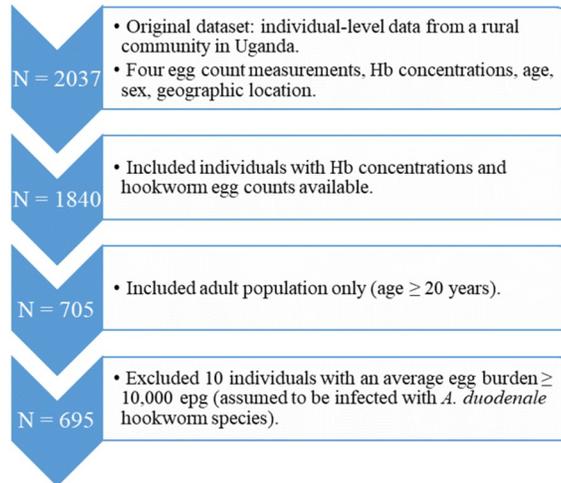
Fig 1. Flowchart describing the process of data selection for the statistical analysis.

https://doi.org/10.1371/journal.pntd.0010279.g001

Model

The association between Hb values and egg counts was studied by means of a hierarchical statistical model defined at individual level, with a latent variable representing the unobserved worm burdens in each human host. The distribution of worm burdens in a population is known to be highly over-dispersed, with most people harbouring no or a small number of worms and only few highly infected individuals. The parasitological part of the hierarchical model is described as follows:

$$w_i | \rho, \delta, \lambda \sim p(w_i | \rho, \delta, \lambda)$$

where w_i is a semi-continuous latent variable representing the worm burden for the i-th individual. Its probability density function is defined as:

$$p(w_i | \rho, \delta, \lambda) = \begin{cases} \rho, & w_i = 0 \\ (1 - \rho) * Weibull(w_i | \delta, \lambda), & w_i > 0 \end{cases}$$

The latent variable w_i was thus modelled from a finite mixture of a Bernoulli distribution with probability ρ (i.e., the probability of having zero worms) and a Weibull distribution with shape parameter δ and scale λ. The employment of this semi-continuous variable is meant to approximate the discrete negative binomial distribution, which has commonly been employed for describing over-dispersed worm burdens in human populations. This approximation was necessary to be able to use the Hamiltonian Monte Carlo algorithm to sample from the posterior distribution, which does not allow the use of a discrete latent variable. The latent variable w_i enabled us to discriminate between infected (if $w_i > 0$) and uninfected individuals (if $w_i = 0$). The corresponding individual egg counts are simulated in the model as follows:

$$x_{i,j} | w_i, ec_i, k_e \sim NB(ec_i, k_e)$$

The variable $x_{i,j}$ is the jth egg count for the i-th individual, with $j = 1,\dots,4$. The parameters ec_i and k_e represent the expectation and the aggregation parameter of the egg count distribution, respectively. The aggregation parameter k_e encapsulates the variability within the same stool specimen and between different stool samples. The expected individual egg burden was assumed to depend on the

Table 1. Parameters employed in the model. The parameters can be either given in input or estimated from the model. For input parameters the corresponding value is reported, the prior distribution for the Bayesian estimation otherwise. When available, the source of the information is cited. Prior distributions are parameterised in terms of mean and standard deviation (normal) or scale (Cauchy).

Parameter	Origin	Value	Prior	Source
Probability of having zero worms ($\rho \in [0, 1]$)	Estimated	-	$\rho \sim \text{Uniform}(0, 1)$	-
Shape of the Weibull distribution for worm burdens in infected individuals ($\delta > 0$)	Estimated	-	$\delta \sim \text{Normal}^+(0.5, 1)$	-
Scale of the Weibull distribution for worm burdens in infected individuals ($\lambda > 0$)	Estimated	-	$\lambda \sim \text{Cauchy}^+(2, 1)$	-
Initial slope of the egg production function (a)	Input	200 epg	-	[12]
Mean saturation level of the egg production function (β)	Input	10,000 epg	-	Pre-set
Shape and rate of the gamma distribution for individual relative susceptibility to infection	Input	50	-	[12]
Aggregation parameter for egg counts ($k_e > 0$)	Estimated	-	$k_e \sim \text{Normal}^+(0, 2)$	[12]
Logarithm of geometric mean of Hb [g/L] distribution in uninfected individuals (μ_0)	Estimated	-	$\mu_0 \sim \text{Normal}(\log(129), 0.1)$	[14,15]
Scale parameter of Hb distribution ($\sigma > 0$)	Estimated	-	$\sigma \sim \text{Normal}^+(0, 0.5)$	[14,15]
Steepness of the logistic decrease ($\varphi_1 > 0$)	Estimated	-	$\varphi_1 \sim \text{Normal}^+(1, 3)$	-
Location of the sigmoid ($\varphi_0 > 0$)	Estimated	-	$\varphi_0 \sim \text{Normal}^+(6, 2)$	-

$^+$ The prior distribution is left-truncated at zero.

https://doi.org/10.1371/journal.pntd.0010279.t001

underlying worm load w_i according to a hyperbolic saturating function [12] describing density dependent fecundity. The function is defined as follows: $ec_i = a * fw_i / \left(1 + \frac{a*fw_i}{b_i}\right)$ where $fw_i = w_i/2$ approximates the female worm load, the parameter a is the average egg production for each female worm in absence of density dependence effects, and b_i the maximum host egg output. The latter was assumed to vary between hosts due to host suitability for infection and was assumed to follow a gamma distribution with mean β representing the population average saturation level in egg production. The values of a and the shape of the gamma distribution (which dictates the level of variation in host suitability) were set to values available from previous work (**Table 1**). The mean saturation level β was set to 10,000 epg to assure the model was able to catch the entire range of observed egg counts.

We further assumed that Hb concentrations (y_i) follow a LogNormal distribution with standard deviation σ representing the natural variation between individuals and a mean that depends on an individual's infection status. In absence of hookworm infection (i.e. $w_i = 0$), we assumed a mean of μ_0. For individuals with a worm burden $w_i > 0$, we assumed the expected Hb to follow a logistic decrease on the logarithmic scale with increasing worm burden, i.e.:

$$y_i | w_i, \mu_0, \sigma, \varphi_0, \varphi_1 \sim LogNormal(\mu_0 - f(w_i), \sigma)$$

where the logistic decrease is defined as $f(w_i) = -\log(1 + e^{\varphi_1 * \log(w_i) - \varphi_0})$. Here, φ_1 represents the steepness of the decreasing function and φ_0 is proportional to the x-axis intercept of the middle point of the sigmoid. The full list of employed parameters is presented in **Table 1**.

The model was fitted to the data within a Bayesian framework. The probability ρ of having zero worms was assigned a Uniform prior between 0 and 1. The hyper-parameters of the Weibull distribution modelling the individual worm burdens in infected individuals were assigned vague hyper-priors. Prior distributions for parameters μ_0 and σ were based on information available from the literature [14,15]. Weakly informative priors were assigned to φ_0, φ_1 and k_e. Prior information on φ_1 and k_e is limited to mathematical constraints.

Posterior distributions of the estimated parameters and predictions were simulated in Stan [16] (code for model specification in Stan is publicly available [17]), a probabilistic programming language for statistical modelling. Stan allows the user to perform full Bayesian statistical inference using Markov chain Monte Carlo (MCMC) sampling (based on NUTS or HMC

sampling algorithm). Simulations were run and further analysed in R (version 3.6.1) [18], using the package *rstan* [19]. The algorithm was run on 4 independent Markov chains, each with 10,000 draws, for a total of 20,000 post-warmup draws (the initial 5,000 draws per chain were considered for warm-up). Convergence was assessed by visual examination of the trace plots and the convergence diagnostic methods provided by Stan. Fit evaluation was performed by means of full posterior predictive checks [20]. In this procedure, the population was replicated for each draw of the joint posterior distribution of model parameters and then compared to the original data.

Simulations

After fitting the model, the posterior distributions obtained for each of the fitted parameters were used to study the appropriateness of the current WHO parasitological target as proxy for hookworm morbidity. We simulated different endemicity scenarios characterised by a specific worm distribution in the population (i.e., a specific combination of the three parameters ρ, δ and λ describing the mixture distribution of worm burdens) resulting in different prevalence of hookworm infection (any intensity and M&HI) settings. The assumptions about the distributions that describe the actual egg counts, the Hb concentrations and the background risk of anaemia in the simulated populations were the same as assumed while fitting the observed data. The simulated infection prevalence ranged from 10% to 90% (any intensity) and 0% to 55% (M&HI). For each scenario, we generated a population of 1000 individuals using a random sample of 5000 draws from the posterior distribution of the parameters and we computed: i) the KK-based hookworm prevalence of any infection, ii) the KK-based prevalence of M&HI infection, iii) overall prevalence of anaemia, iv) anaemia attributable to hookworm infection. The latter is defined as the difference between the overall predicted anaemia and a baseline anaemia in the counterfactual scenario without worm infections, estimated from the model. This computation does not explore the contribution of hookworm infection to the severity of anaemia. Each of the aforementioned quantities were computed based on 3 different sampling schemes to assess the KK-based prevalence of hookworm: i) a single KK slide, which is the most common diagnostic scheme in the field, ii) duplicate slides, and iii) four repeated slides, in line with the dataset used in this study.

All analyses were performed in accordance with the Policy-Relevant Items for Reporting Models in Epidemiology of Neglected Tropical Diseases (PRIME-NTD) criteria [21] (**S1 Table**). The full model specification and the code for data selection, statistical analysis and simulations are publicly available [17]. A flow diagram summarising the complete process of model specification, fitting and simulation is provided as supporting material (**S1 Fig**).

Results

Association between individual haemoglobin levels and egg counts

Our statistical model fitted the individual data from Uganda well, which suggests that a logistic function is a reasonable representation of the effect of increasing hookworm burden on Hb levels in adults (**Fig 2**).

This figure shows the joint posterior predictive distribution of Hb and egg counts based on 20,000 iterations, displayed via the mean of Hb predictions (y-axis) and the geometric mean of egg count predictions (x-axis). The 95% Bayesian Credible Interval (BCI) is solely displayed for the marginal Hb posterior predictive distribution. In absence of hookworm infection, Hb levels in the population are distributed with an estimated geometric mean of 126 g/L (95%-BCI: 124.3 g/L– 127.5 g/L) and a standard deviation of the Hb natural logarithm equal to 0.16 (95%-BCI: 0.15–0.17). The estimated parameters of the logistic function showed that

hookworm infection has little effect on Hb levels in the population for intensities up to 1000 epg and that Hb falls below the WHO threshold for anaemia when intensities are above 2000 epg. See **Table 2** for detailed parameter estimates.

The posterior predictive checks obtained from the model fit to the data recorded a prevalence of hookworm infection in the population of 54% (95%-BCI: 48%-59%; 52% in data), a M&HI infection prevalence of 4.5% (95%-BCI: 3% - 7%; 4% in data) and a prevalence of overall anaemia of 40% (95%-BCI: 36% - 45%; 41% in data). The posterior distributions of the parameters μ_0 and σ suggested a baseline prevalence of anaemia (in absence of hookworm infection) in the population of 38% (95%-BCI: 35% - 41%). This represents the prevalence of anaemia that may be attributed to other causes.

Full posterior predictive checks for Hb levels and egg counts distributions, based on 20,000 replicated datasets were compared to the original data and used to assess model fit (**Fig 3**). The posterior predictive distributions approximated the observed egg counts and Hb distributions in the population reasonably well.
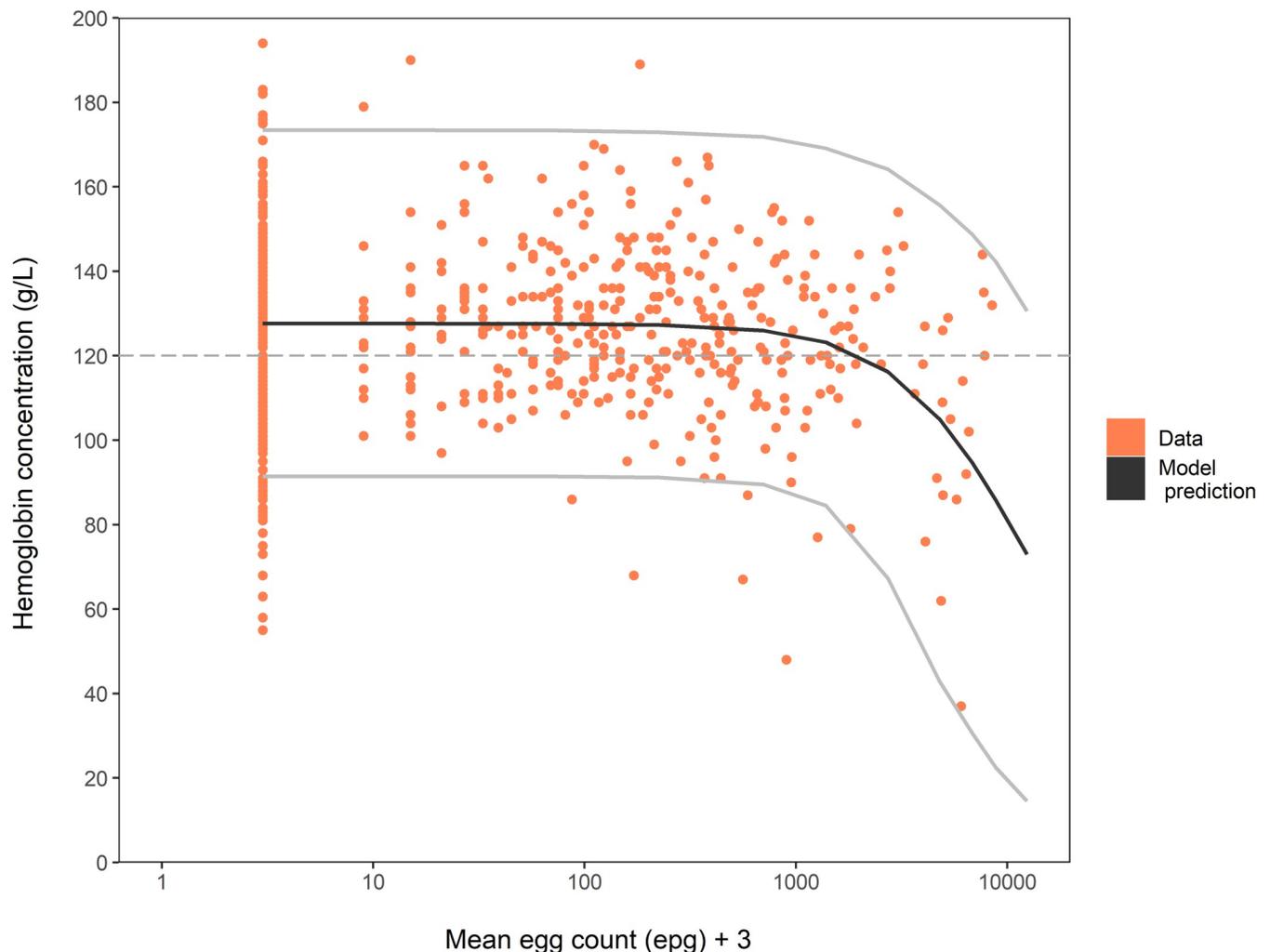
Fig 2. Relationship between individual haemoglobin concentrations (g/L) and mean egg count (epg). Orange bullets represent observed individual data used for model fit. The joint posterior predictive distribution of Hb and mean egg count based on 20,000 replications is displayed through the mean of the Hb predictions (y-axis) and the geometric mean of egg count predictions (x-axis) (black line). Mean egg count data are shifted of +3 epg (i.e. half of the detection limit) to display the zero counts on the logarithmic-scaled axes. The 95% Bayesian Credible Interval (light grey lines) is shown for the marginal Hb posterior predictive distribution. The WHO threshold defining anaemia used in the model is displayed as horizontal dashed line.
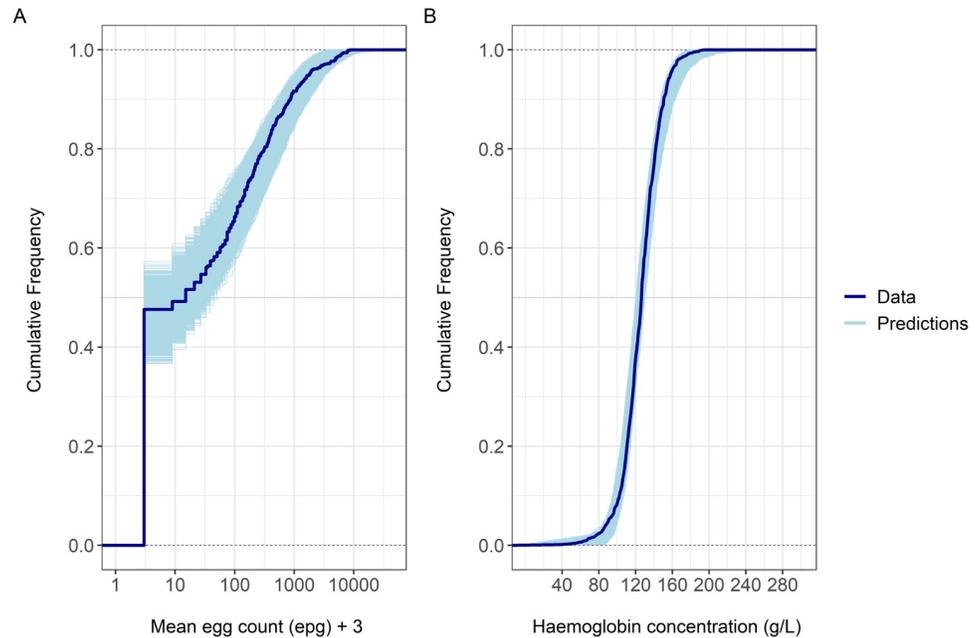
Fig 3. Comparison of observed data and model prediction. (A) Full posterior predictive checks for the mean egg counts. Mean egg count data are shifted of +3 epg (i.e. half of the detection limit) to display the zero counts on the logarithmic-scaled axes. (B) Full posterior predictive checks for haemoglobin concentrations. Dark lines represent the corresponding distribution observed in the Ugandan population. Both marginal predictive distributions are displayed through the cumulative distribution function of the predictions based on 20,000 replications (light lines) and compared with observed data.

https://doi.org/10.1371/journal.pntd.0010279.g003

Simulated prevalence of anaemia due to hookworm infection

The posterior distributions obtained for each of the fitted parameters were used for simulating different endemicity scenarios, assuming that the risk of anaemia due to causes other than hookworm infection is the same as in the Ugandan data. The simulated prevalence of anaemia attributable to hookworm is displayed in **Fig 4** with its mean and corresponding 95% prediction interval (PI), as a function of infection prevalence (any intensity and M&HI). The figure shows that the prevalence of anaemia attributable to hookworm ranged from 0% to 25% (95%-PI: 20% - 30%) by varying endemicity scenarios and it showed strong association with the prevalence of M&HI

Table 2. Model parameter estimates. Estimates represent posterior distributions of model parameters. For each parameter, Rhat can be used as indicator for Markov chains convergence (at convergence, Rhat = 1). BCI = Bayesian credible interval, based on the central 95% percentiles of the posterior draws.

Parameter	Mean [95%-BCI]	Rhat
Probability of having zero worms ($\rho \in [0, 1]$)	0.38 [0.32, 0.43]	1.001
Shape of the Weibull distribution for worm burdens in infected individuals ($\delta > 0$)	0.48 [0.40, 0.55]	1.001
Scale of the Weibull distribution for worm burdens in infected individuals ($\lambda > 0$)	3.31 [2.34, 4.41]	1.001
Aggregation parameter for egg counts ($k_e > 0$)	0.66 [0.59, 0.74]	1.000
Logarithm of geometric mean of Hb [g/L] distribution in uninfected individuals (μ_0)	4.84 [4.82, 4.85]	1.000
Scale parameter of Hb distribution ($\sigma > 0$)	0.16 [0.15, 0.17]	1.000
Steepness of the logistic decrease ($\varphi_1 > 0$)	1.90 [1.33, 2.54]	1.001
Location of the sigmoid ($\varphi_0 > 0$)	9.23 [6.96, 11.90]	1.000

https://doi.org/10.1371/journal.pntd.0010279.t002

infections. In all the simulated scenarios, a 2% prevalence of M&HI infections in adults corresponded to a prevalence of anaemia attributable to hookworm infections lower than 1%.

Our simulations showed that the prevalence of infection is sensitive to the number of KK simulated measurements. However, low values of M&HI infection prevalence (<15%) in a given scenario were relatively consistent over the three sampling schemes (1, 2 or 4 KK simulated measurements).

The endemicity scenarios presented in Fig 4 are based on a fixed value of the shape parameter for the distribution of worms across infected individuals (δ = 0.5, i.e., approximately the value estimated for the Ugandan dataset). Simulations for alternative values of the shape parameter resulted in a similar pattern, with the prevalence of anaemia attributable to hookworm ranging from 0% to 30% (95%-PI: 24% - 36%), by varying endemicity scenarios (S2 Fig).
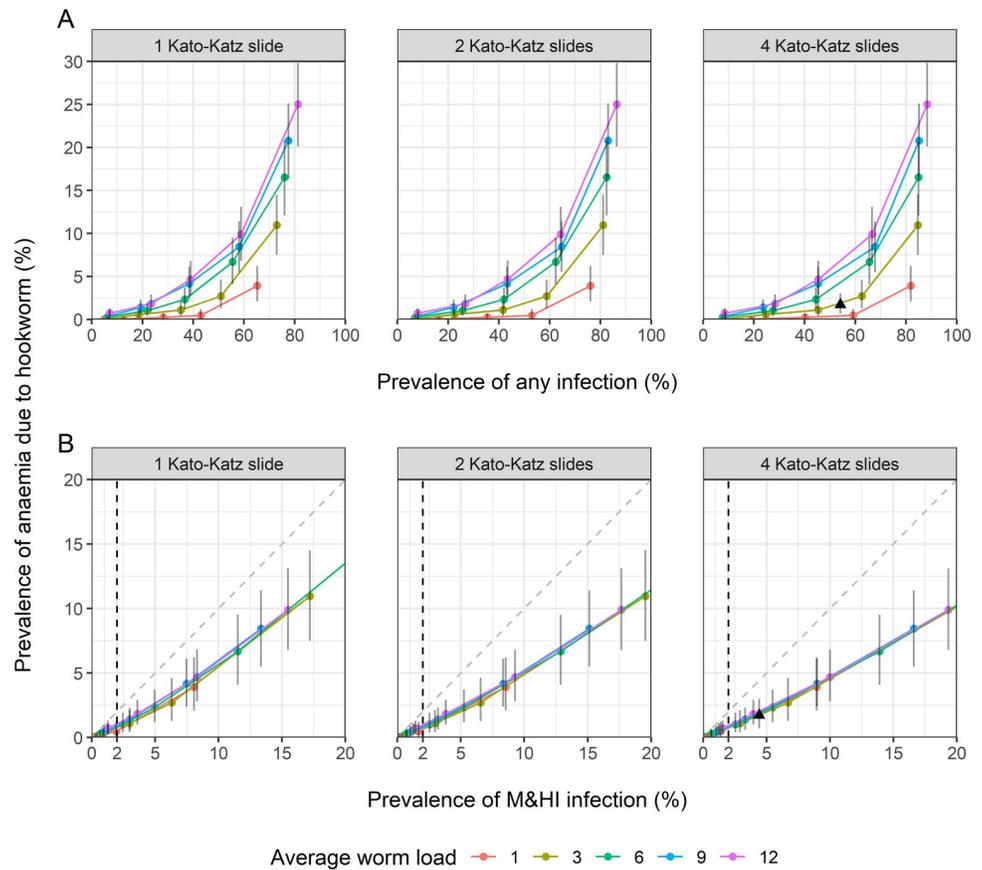
Fig 4. Estimated prevalence of anaemia attributable to hookworm in adult populations, for different infection prevalence scenarios. Predictions are displayed for different infection prevalence scenarios, defined by varying the probability ρ of having zero worms among the five values 0.1, 0.3, 0.5, 0.7, 0.9 (different bullets of the same colour) and the population mean worm burden (different colours). The shape parameter of the worm distribution across infected individuals is fixed to δ = 0.5 (as estimated from the Ugandan data). From left to right, the prevalence of any hookworm infection (A) and moderate-to-heavy intensity of infection (B) is based on 1, 2, 4 Kato-Katz slides, respectively. Each point (x, y) represents the mean predictions of infection prevalence (x-value) vs. hookworm-anaemia prevalence (y-value). The error band reports the 95% prediction interval (PI) of predicted hookworm-anaemia. The black triangle indicates the predictions (mean and 95%-PI) for the scenario resulting from the fit to the Ugandan dataset. Dashed vertical lines mark the WHO morbidity threshold of 2% moderate-to-heavy intensity prevalence. The bisector in panels (B) is added to visualise the robustness of the prevalence of moderate-to-heavy intensity of infection over the three sampling schemes.

Discussion

The aim of this study was to estimate the prevalence of anaemia attributable to hookworm infection and to evaluate whether the current guidelines for hookworm control properly reflect actual hookworm morbidity. To that end, we used a statistical hierarchical model that estimates the mechanisms of Hb decrease in response to increasing worm burden at individual level. Our analysis show that the estimated hookworm-attributable anaemia is strongly associated to M&HI infections in adults. The current threshold used as an indicator for the elimination of hookworm morbidity (2% of M&HI prevalence in school-aged children as for the 2030 WHO target) corresponds to less than 1% prevalence of anaemia due to hookworm in the simulated endemicity settings. Therefore, our predictions suggest that the current threshold adopted by the WHO for the elimination of hookworm as a public health problem seems suitable. It should be noted that our results are solely based on Hb values and egg counts from adults, due to limited available data. Even though children show a higher age/sex-variation in the definition of the anaemia thresholds and the underlying Hb distribution, we would expect a similar association between the prevalence of anaemia attributable to hookworm and the prevalence of hookworm infection in adults as in children. In fact, it is known that adults can be infected with hookworm to the same extent as children [22] or even more, with a higher experienced morbidity [23]. In addition, our analysis confirms findings from a previous study in a population of school-aged children in Zanzibar [24] which showed an approximately linear association of intestinal blood loss with intensity of hookworm infection.

Hookworm infections were found to have little effect on Hb levels for intensities up to 1000 epg. In fact, it is known that the uptake of blood by few worms should be easily counterbalanced by stored iron in the host, with some variation depending on individual nutritional reserves [3]. The development of anaemia as a consequence of hookworm infection depends on individual iron balance. At a population level, the threshold in the relationship between intensity of hookworm infection and prevalence of anaemia depends on the underlying iron reserves of the population [24,25], here incorporated as other sources of anaemia than hookworm infection. Our analysis show that for moderate ($\geq$ 2000 epg) and heavy ($\geq$ 4000 epg) intensity of infection the predicted average Hb drastically drops below the WHO threshold for anaemia. These results are in line with Lwambo *et al.* [9].

The simulated scenarios also show that the predicted KK based prevalence is dependent on the number of repeated measurements, but values of M&HI infection prevalence < 15% are quite robust over the three simulated sampling schemes (1, 2 or 4 KK). Recent work by our team [26] reached similar conclusions, suggesting that 1 slide/sample may be sufficient for determining prevalence of M&HI infection to assess the morbidity target.

Previous modelling studies estimated the prevalence of hookworm morbidity from existing data [3,9,27] without explicitly considering anaemia due to other causes. In our model we accounted for an underlying prevalence of anaemia due to other causes such as poor nutritional status or suffering from other infectious diseases, by explicitly modelling individual Hb variation. The latter is mainly governed by the scale parameter σ of the Hb distribution, defined to be equal for both groups of infected and not infected individuals, thus reflecting an underlying prevalence of anaemia due to other causes in the total population. This approach finds support from a recently published systematic review and meta-analysis [28] which showed that the intensity of hookworm infection and the influence of other sources on the risk of anaemia drive differences in mean Hb concentrations, in hookworm endemic settings.

The Ugandan dataset used to fit our model showed some peculiarities we had to account for in our analysis. The subset of population detected with zero eggs for hookworm reported relatively low average Hb, presumably indicating other abundant sources of anaemia: part of

the population was co-infected with *Schistosoma mansoni* [10]. The majority of individuals who tested positive to hookworm was lightly infected (only 4% of M&HI infection prevalence). Therefore, we did not include sex differences in the model, with the aim of gathering enough information about the steepness of the Hb decrease. However, we expect male and female population to show the same decreasing dynamics in response to increasing worm burdens, considering biological differences in how Hb concentrations are distributed in absence of infection between the two groups. The analysed Ugandan population is also heterogeneous with respect to the hookworm species causing the infection. In light of considerable differences between hookworm species in egg production [8], density-dependent fecundity effects [29] and in contribution to morbidity [3,13,25], we excluded ten individuals that we assumed to be infected with *A. duodenale* (detected epg $\geq$ 10,000). The differentiation of hookworm species in the present model framework would require a species-specific parametrisation of the parasitological dynamics, which is beyond the information available in the analysed data. Future versions of our statistical model may allow for different species, if sufficiently detailed data of species-specific egg counts (e.g., using qPCR [30]) and haemoglobin become available.

Few individuals infected with *A. duodenale* at lower intensities could not be identified and could well still be part of the analysed data, contributing to a milder estimated slope of the Hb logistic decrease. We nevertheless do not expect those data to have a critical impact on the overall estimation of the relationship between Hb and worm burdens.

The distribution of worms between individuals is commonly described by a negative binomial distribution, parameterised with a mean worm burden and an aggregation parameter regulating the level of over-dispersion. However, we coded the model in Stan [16] which does not allow the use of a discrete latent variable [16]. Therefore, we used a semi-continuous approximation resulting from a finite mixture of a Weibull distribution and a point mass at zero, that can be interpreted as the level of true infection intensity in the host. The employment of a finite mixture distribution for modelling worm burdens allows discrimination between infected and uninfected individuals, providing a prediction of the true prevalence of infection (i.e., presence of worms). In our model we account for possible false negative results from the KK test, that we assume to be ascribed to an undetectable level of infection (low worm loads) or to the sensitivity of the test (simulated through the over-dispersion parameter of the negative binomial distribution used for modelling egg counts).

In conclusion, this study represents a first step towards better quantifying the prevalence of anaemia due to hookworm infection. Our results support the appropriateness of the current WHO threshold of 2% prevalence of M&HI as a proxy for hookworm morbidity. In addition, a single KK slide may be sufficient for assessing this morbidity target. Further studies are needed to elucidate Hb dynamics pre- and post- mass drug administration, where the infection harbouring time plays an important role [28], ideally using longitudinal data collected in both children and adults.

Supporting information

S1 Fig. Flow diagram summarising the complete process of modelling, fitting and simulation of the statistical analysis for hookworm anaemia.
(TIF)

S2 Fig. Estimated prevalence of anaemia attributable to hookworm in adult populations, for different infection prevalence scenarios. Predictions are displayed for different infection prevalence scenarios, defined by varying the probability ρ of having zero worms among the five values 0.1, 0.3, 0.5, 0.7, 0.9 (different points of the same colour), the population mean worm burden (different colours) and the shape parameter δ of worm distribution in infected

individuals (different shapes). From left to right, the prevalence of any hookworm infection (A) and moderate-to-heavy intensity of infection (B) is based on 1, 2, 4 Kato-Katz slides, respectively. Each point (x, y) represents the mean predictions of infection prevalence (x-value) vs. hookworm-anaemia prevalence (y-value). The error band reports the 95% Confidence Interval (CI) of predicted hookworm-anaemia. Dashed vertical lines mark the WHO morbidity threshold of 2% moderate-to-heavy intensity prevalence. The bisector in panels (B) is added to visualise the robustness of the prevalence of moderate-to-heavy intensity of infection over the three sampling schemes.
(TIF)

S1 Table. The Policy-Relevant Items for Reporting Models in Epidemiology of Neglected Tropical Diseases (PRIME-NTD).
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Veronica Malizia, Federica Giardina, Sake J. de Vlas, Luc E. Coffeng.

Data curation: Veronica Malizia, Luc E. Coffeng.

Formal analysis: Veronica Malizia, Federica Giardina, Luc E. Coffeng.

Funding acquisition: Sake J. de Vlas, Luc E. Coffeng.

Methodology: Veronica Malizia, Federica Giardina, Luc E. Coffeng.

Software: Veronica Malizia.

Supervision: Federica Giardina, Sake J. de Vlas, Luc E. Coffeng.

Validation: Veronica Malizia, Federica Giardina, Luc E. Coffeng.

Visualization: Veronica Malizia, Federica Giardina, Luc E. Coffeng.

Writing – original draft: Veronica Malizia, Federica Giardina.

Writing – review & editing: Veronica Malizia, Federica Giardina, Sake J. de Vlas, Luc E. Coffeng.

References

1. World Health Organization. Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups. 2017.

2. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit Vectors. 2014; 7:37. https://doi.org/10.1186/1756-3305-7-37 PMID: 24447578

3. Crompton DW, Whitehead RR. Hookworm infections and human iron metabolism. Parasitology. 1993; 107(Suppl 1):S137–S145. https://doi.org/10.1017/s0031182000075569 PMID: 8115178

4. Kassebaum NJ. The global burden of anemia. Hematology/Oncology Clinics. 2016; 30(2):247–308. https://doi.org/10.1016/j.hoc.2015.11.002 PMID: 27040955

5. Montresor A, Crompton DWT, Hall A, Bundy DA, Savioli L, World Health O. Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level: a guide for managers of control programmes. World Health Organization, 1998.

6. World Health Organization. 2030 targets for soil-transmitted helminthiases control programmes. 2020.

7. Smith JL, Brooker S. Impact of hookworm infection and deworming on anaemia in non-pregnant populations: a systematic review. Trop Med Int Health. 2010; 15(7):776–795. https://doi.org/10.1111/j.1365-3156.2010.02542.x PMID: 20500563

8. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet. 2006; 367(9521):1521–1532. https://doi.org/10.1016/S0140-6736(06)68653-4 PMID: 16679166

9. Lwambo NJ, Bundy DA, Medley GF. A new approach to morbidity risk assessment in hookworm endemic communities. Epidemiol Infect. 1992; 108(3):469–481. https://doi.org/10.1017/s0950268800049980 PMID: 1601081

10. Pullan RL, Kabatereine NB, Quinnell RJ, Brooker S. Spatial and genetic epidemiology of hookworm in a rural community in Uganda. PLoS Negl Trop Dis. 2010; 4(6):e713. https://doi.org/10.1371/journal.pntd.0000713 PMID: 20559556

11. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. World Health Organization. 2011;WHO/NMH/NHD/MNM/11.1.

12. Coffeng LE, Bakker R, Montresor A, de Vlas SJ. Feasibility of controlling hookworm infection through preventive chemotherapy: a simulation study using the individual-based WORMSIM modelling framework. Parasites & Vectors. 2015; 8(1):541. https://doi.org/10.1186/s13071-015-1151-4 PMID: 26489659

13. Jonker FA, Calis JC, Phiri K, Brienen EA, Khoffi H, Brabin BJ, et al. Real-time PCR demonstrates Ancylostoma duodenale is a key factor in the etiology of severe anemia and iron deficiency in Malawian preschool children. PLoS neglected tropical diseases. 2012; 6(3):e1555. https://doi.org/10.1371/journal.pntd.0001555 PMID: 22514750

14. Pullum T, Collison DK, Namaste S, Garrett D. Hemoglobin data in DHS Surveys: Intrinsic variation and measurement error. Rockville, Maryland, USA: ICF, 2017.

15. Kassebaum NJ, Jasrasaria R, Naghavi M, Wulf SK, Johns N, Lozano R, et al. A systematic analysis of global anemia burden from 1990 to 2010. Blood. 2014; 123(5):615–624. https://doi.org/10.1182/blood-2013-06-508325 PMID: 24297872

16. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.19.1. 2019. Available from: https://mc-stan.org.

17. Malizia V. Hookworm anaemia model. Available from: https://github.com/VeronicaMalizia/Hookworm_anaemia2021.

18. Team RC. R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/.

19. Stan Development Team. RStan: the R interface to Stan. R package version 2.21.2. 2020. Available from: http://mc-stan.org/.

20. Green MJ, Medley GF, Browne WJ. Use of posterior predictive assessments to evaluate model fit in multilevel logistic regression. Vet Res. 2009; 40(4):30. https://doi.org/10.1051/vetres/2009013 PMID: 19323968

21. Behrend MR, Basanez MG, Hamley JID, Porco TC, Stolk WA, Walker M, et al. Modelling for policy: The five principles of the Neglected Tropical Diseases Modelling Consortium. PLoS Negl Trop Dis. 2020; 14 (4):e0008033. https://doi.org/10.1371/journal.pntd.0008033 PMID: 32271755

22. Truscott JE, Ower AK, Werkman M, Halliday K, Oswald WE, Gichuki PM, et al. Heterogeneity in transmission parameters of hookworm infection within the baseline data from the TUMIKIA study in Kenya. Parasites & Vectors. 2019; 12(1):442. https://doi.org/10.1186/s13071-019-3686-2 PMID: 31522687

23. Brooker S, Bethony J, Hotez PJ. Human hookworm infection in the 21st century. Adv Parasitol. 2004; 58:197–288. https://doi.org/10.1016/S0065-308X(04)58004-1 PMID: 15603764

24. Stoltzfus RJ, Albonico M, Chwaya HM, Savioli L, Tielsch J, Schulze K, et al. Hemoquant determination of hookworm-related blood loss and its role in iron deficiency in African children. Am J Trop Med Hyg. 1996; 55(4):399–404. https://doi.org/10.4269/ajtmh.1996.55.399 PMID: 8916795

25. Roche M, Layrisse M. The nature and causes of" hookworm anaemia.". American Journal of Tropical Medicine and Hygiene. 1966; 15(6, Pt. 2):1031–1102.

26.   Coffeng LE, Malizia V, Vegvari C, Cools P, Halliday KE, Levecke B, et al. Impact of Different Sampling Schemes for Decision Making in Soil-Transmitted Helminthiasis Control Programs. J Infect Dis. 2020; 221(Suppl 5):S531–S538. https://doi.org/10.1093/infdis/jiz535 PMID: 31829425

27.   Chan MS, Medley GF, Jamison D, Bundy DA. The evaluation of potential global morbidity attributable to intestinal nematode infections. Parasitology. 1994; 109 (Pt 3):373–387. https://doi.org/10.1017/s0031182000078410 PMID: 7970892

28.   Byrne A, Manalo G, Clarke NE, Vaz Nery S. Impact of hookworm infection and preventive chemotherapy on haemoglobin in non-pregnant populations. Trop Med Int Health. 2021; 26(12):1568–1592. https://doi.org/10.1111/tmi.13681 PMID: 34587315

29.   Anderson RM, Schad GA. Hookworm burdens and faecal egg counts: an analysis of the biological basis of variation. Trans R Soc Trop Med Hyg. 1985; 79(6):812–825. https://doi.org/10.1016/0035-9203(85)90128-2 PMID: 3832493

30.   Cools P, Vlaminck J, Verweij JJ, Levecke B. Quantitative PCR in soil-transmitted helminth epidemiology and control programs: Toward a universal standard. PLoS neglected tropical diseases. 2021; 15(3): e0009134. https://doi.org/10.1371/journal.pntd.0009134 PMID: 33661910