A general framework to support cost-efficient survey design choices for the control of soil-transmitted helminths when deploying Kato-Katz thick smear

Running title: Survey designs for STH control programs

Adama Kazienga\textsuperscript{1,2*}, Bruno Levecke\textsuperscript{1}, Gemechu Tadesse Leta\textsuperscript{3}, Sake J. de Vlas\textsuperscript{2}, Luc E. Coffeng\textsuperscript{2}

\textsuperscript{1}Department of Translational Physiology, Infectiology and Public Health, Ghent University, Merelbeke, Belgium

\textsuperscript{2}Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

\textsuperscript{3}Bacterial, Parasitic and Zoonotic Diseases Research Directorate, Ethiopian Public Health Institute, Addis Ababa, Ethiopia

*Corresponding author: kazienga_adama@yahoo.fr (KA).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background

To monitor and evaluate soil-transmitted helminth (STH) control programs, the World Health Organization (WHO) recommends screening stools from 250 children across 5 schools, deploying Kato-Katz thick smear (KK). However, it remains unclear whether these recommendations are sufficient to make adequate decisions about stopping preventive chemotherapy (PC) (prevalence of infection <2%) or declaring elimination of STH as a public health problem (prevalence of moderate-to-heavy intensity (MHI) infections <2%).

Methodology

We developed a simulation framework to determine the effectiveness and cost of survey designs for decision-making in STH control programs, capturing the operational resources to perform surveys, the variation in egg counts across STH species, across schools, between and within individuals, and between repeated smears. Using this framework and a lot quality assurance sampling approach, we determined the most cost-efficient survey designs (number of schools, subjects, stool samples per subject, and smears per stool sample) for decision-making.

Principal findings

For all species, employing duplicate KK (sampling 4 to 6 schools and 64 to 70 subjects per school) was the most cost-efficient survey design to assess whether prevalence of any infection intensity was above or under 2%. For prevalence of MHI infections, single KK was the most cost-efficient (sampling 11 to 25 schools and 52 to 84 children per school).
Conclusions/Significance

KK is valuable for monitoring and evaluation of STH control programs, though we recommend to deploy a duplicate KK on a single stool sample to stop PC, and a single KK to declare the elimination of STH as a public health problem.

Keywords: soil-transmitted helminths, Kato-Katz thick smear, lot quality assurance sampling, survey design, total operational cost
Author summary

Worldwide large-scale deworming programs are implemented to reduce the morbidity attributable to intestinal worms in school children. To monitor and evaluate the progress towards the programs goals, World Health Organization (WHO) has developed both a survey design and a corresponding decision tree based on the prevalence and intensity of infections. However, these programs operate in resource-constrained countries, and hence it is crucial to minimize the operational costs to survey worm infections while ensuring the correctness of the program decision. To further support WHO in more evidence-based recommendations for cost-efficient decision-making, we developed a general framework that captures both the operational resources to perform surveys and the variation in test results when deploying the current diagnostic standard. Subsequently, we determined the most cost-efficient survey design to decide to stop the deworming programs and to verify whether the morbidity attributable to intestinal worms has been eliminated as a public health problem. Generally, we found that the current WHO-recommended survey design may not allow for optimal decision making. Based on our results, we proposed alternative survey designs for each of the worm species and program targets.
Introduction

Recently, the World Health Organization (WHO) released its 2030 roadmap for the control of soil-transmitted helminths (STHs) [1,2]. In this new roadmap, the two crucial targets are (i) to eliminate STHs as a public health problem (EPHP), defined as prevalence of moderate-to-heavy intensity (MHI) infections less than 2%, and (ii) to reduce the number of tablets needed in preventive chemotherapy (PC). Regarding this second target, WHO recommends to stop PC if the prevalence of infection (any intensity) is less than 2%.

To assess whether prevalence of any intensity or MHI infection is under 2%, the WHO recommends screening stool samples from 250 children across 5 schools (50 children per school), using single Kato-Katz thick smears (KK) [2]. However, little attention has been paid to whether this survey design is sufficient for reliable decision-making. There will always be a risk of making a wrong decision: stopping PC or declaring EPHP too early (referred to as “undertreatment” from here on) or too late (“overtreatment”). Amongst other things, these risks are driven by survey sample size and diagnostic accuracy, which means that a trade-off has to be made between operational costs and feasibility of particular survey designs, and the maximum acceptable risks of under- and overtreatment.

Here, we assess the trade-off of survey design and associated operational costs versus the risk of under- and overtreatment for STHs when using KK as a diagnostic method because it was found more cost-effective than alternative diagnostic techniques based on egg counting [3]. We do this using a previously developed simulation framework for lot quality assurance sampling (LQAS) that only captured the
prevalence of infection and not the variation in intensity of infection [4]. For the purpose of this study, we expanded this framework to also capture variation in intensity of infection across STH species (Ascaris lumbricoides, Trichuris trichiura, hookworms), across schools, between and within individuals, and between repeated thick smears. These sources of variation are important as the sensitivity of KK for detecting infection in an individual depends on the intensity of infection [5–9]. Furthermore, using this expanded framework, we assess to what extent the accuracy of survey results can be improved (and at what cost) by increasing the number of thick smears per stool sample or the number of stool samples per person [5,10]. Finally, for different levels of acceptable risk of under- and overtreatment, we determine the most cost-efficient survey design and the associated decision cut-off (maximum number of egg-positive individuals) for decision-making.

Methods

Overview

Based on a previously developed LQAS framework for monitoring and evaluation (M&E) of neglected tropical diseases (NTDs) [4], we developed a generic simulation framework for STH surveys and the associated operational costs at the level of a PC implementation unit. Using this framework, we assessed the cost-efficiency of different survey designs (number of schools, number of children per school, number of stool samples per child, and number of KK thick smears per stool sample) for each STH species and decision type (stop PC or declare EPHP) via the following steps: (1) simulate egg counts in thick smears of sampled children and determine the number of children that test positive; (2) for all possible decision cut-offs in terms of maximum
number of positive children, calculate the probability of overtreatment and undertreatment across repeated Monte Carlo simulations and determine the decision cut-off for adequate decision-making (acceptable risk of under- and overtreatment); (3) estimate the total survey cost. To determine the most cost-efficient survey design and associated cut-off, we selected all designs that resulted in adequate decision-making and compared them in terms of cost and feasibility to implement.

**Simulation of egg counts for different survey designs**

We adopted the simulation framework for egg counts developed by Coffeng et al. [11] and expanded it with geographical variation in infection levels. The expanded framework simulates egg count data from a compound lognormal-gamma-gamma-gamma-Poisson distribution, capturing the following sources of variability in egg counts:

1. Variability in mean egg per gram stool (EPG) between schools (assumed to be lognormal distributed (see **Supplement S1**));
2. Inter-individual variability in EPG due to variation in infection levels between individuals (considered to be gamma-distributed), where the level of aggregation (shape of the gamma distribution) is a linear function of the school-level mean EPG (see **Supplement S1**);
3. Day-to-day variability in mean EPG within an individual due to heterogeneous egg shedding over time (assumed to be gamma-distributed);
4. Variability in mean EPG between repeated slides based on the same sample due to the aggregated distribution of eggs in feces (assumed to follow a gamma distribution);
5. Variability in count observation (expressed in raw egg count) due to random diagnostic variation (assumed to be Poisson distributed).

We presented the mathematical backbone of the simulation framework in *Supplement S2* and quantified these aforementioned sources of variation based on published datasets [5,12]. Table 1 provides the full overview of the estimated parameter values for each STH species.

**Table 1.** Parametrization of the simulation framework for various sources of variability in Kato-Katz thick smear egg counts. We parametrized the lognormal distribution using the mean and standard deviation on the logarithmic scale, and for the gamma distribution, the shape parameter $k$ and rate ($\mu_k$), where $\mu$ is the distribution’s mean. To quantify $k$, the coefficient of variation ($cv$) was used as standardized measure of variability [13].

<table>
<thead>
<tr>
<th>Parameters</th>
<th>STH species</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variability in mean EPG across schools (within the same district ($\sigma_i$))</td>
<td>Ascaris: 0.69, Hookworm: 0.80, Trichuris: 0.52</td>
<td>[12]</td>
</tr>
<tr>
<td>Intercept for school-level aggregation ($\beta_0$) as a linear function of the school-level mean EPG</td>
<td>Ascaris: 0.0158, Hookworm: 0.0162, Trichuris: 0.0098</td>
<td>[12]</td>
</tr>
<tr>
<td>Slope for school-level aggregation ($\beta_1$) as a linear function of the school-level mean EPG</td>
<td>Ascaris: 0.0019, Hookworm: 0.0222, Trichuris: 0.0444</td>
<td>[12]</td>
</tr>
<tr>
<td>Day-to-day variation in EPG within an individual (shape $k_d$, expressed as $\frac{1}{cv^2}$)</td>
<td>Ascaris: 0.5102, Hookworm: 0.8734, Trichuris: 1.4172</td>
<td>[3]</td>
</tr>
</tbody>
</table>
Using this simulation framework, we considered 4 survey designs, denoted as $KK_a \times b$.

$a$ is the number of stool samples per person, and $b$ is the number of repeated smears per stool sample. These survey designs were $KK_1 \times 1$ (considered as the reference [1,2]), $KK_1 \times 2$, $KK_2 \times 1$, and $KK_2 \times 2$. We further set the number of schools ($n_{schools}$) ranging from 3 to 10 for any intensity of infection and from 10 to 25 for MHI infections, while the number of children per school ($n_{children}$) ranged from 10 to 200 for both targets.

Calculating the probability of under- and overtreatment

To determine the probability of making right and wrong policy decisions for different survey designs, we used a 2-stage LQAS approach, a framework for NTD control program decision-making which we have described in detail previously [4]. The 2-stage aspect refers to the fact that we explicitly consider that survey results will be sampled from multiple clusters. In 2-stage LQAS, decision-making is based on the total number of positive test results ($X_+$; here, egg-positive individuals) in a sample of pre-defined size from a pre-defined number of clusters in an implementation unit and a decision cut-off $c$. The frequency of PC will be reduced (or PC is stopped altogether) in case $X_+$ is less than the decision cut-off $c$. Otherwise, the frequency of PC remains unchanged or may even be increased. This framework was used to assess two features of the decision: (1) the operational program decision threshold (e.g., 2% prevalence of infection); (2) the maximum acceptable risk of making a wrong decision when the true prevalence based on a single KK is at a pre-defined lower or higher value than the operational decision threshold (i.e., the “grey zone”, e.g., 1% and 3%).
Using LQAS, we assessed two operational decision thresholds for each STH species: one for a 2% prevalence of any intensity of infection (i.e., related to stopping PC) and another for a 2% prevalence of MHI infections (i.e., related to declare EPHP). Then, we considered adequate decision-making by setting the maximum allowed probability of unnecessarily continuing with the program ($E_{\text{overtreat}}$) when true prevalence is 3% (the upper limit of the grey zone, or $U_L$) at 0.25. We further set the highest allowed probability of prematurely stopping interventions ($E_{\text{under treat}}$) when true prevalence is 1% (the lower limit of the grey zone, or $L_L$) at 0.05. For this, we determined the mean EPG at the implementation unit level that corresponds to these limits ($L_L$ and $U_L$) as measured by single KK (Table 2), conditional on the estimated parameter values for sources of variability in egg counts (Table 1), assuming 100% specificity of KK [14,15].

Table 2. Parametrization of simulations for 2-stage lot quality assurance sampling in terms of the natural logarithm of the geometric mean infection intensity (eggs per gram feces or EPG) in the population as measured by single Kato-Katz.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Ascaris</th>
<th>Hookworm</th>
<th>Trichuris</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Any intensity infections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower limit mean EPG</td>
<td>log(0.449)</td>
<td>log(0.280)</td>
<td>log(0.298)</td>
</tr>
<tr>
<td>Upper limit mean EPG</td>
<td>log(6.91)</td>
<td>log(1.013)</td>
<td>log(0.980)</td>
</tr>
<tr>
<td><strong>Moderate-to-heavy intensity infections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower limit mean EPG</td>
<td>log(314)</td>
<td>log(164)</td>
<td>log(145)</td>
</tr>
<tr>
<td>Upper limit mean EPG</td>
<td>log(620)</td>
<td>log(287)</td>
<td>log(217)</td>
</tr>
</tbody>
</table>
Calculating the operational costs and determining the most cost-efficient survey design

We estimated the total survey costs for each simulated survey design and STH species. This cost was composed of (i) the cost of consumables to collect and process samples, (ii) the personnel costs, and (iii) the travel costs [4]. Technical details on how the total survey costs were estimated can be found in Supplement S3.

To determine the most cost-efficient survey design and associated decision cut-off, we selected all designs that allowed for adequate decision-making. Then, we chose the survey design that resulted in the lowest total survey cost ($C_{tot}$) and required a maximum of 100 children per school (a reasonable maximum size of individuals in a school).

Sensitivity analysis

We assessed the impact of several critical parameters on the estimated cost-efficiency of different survey designs. First, we explored the impact of assuming that variation between individuals within schools is governed by a fixed aggregation parameter $k_k$ (in contrast to letting it vary as a linear function of the mean EPG in a school). Second, we assessed the impact of considering higher inter-individual variability in egg counts ($k_k$) in (post-) control scenarios due to incomplete deworming coverage compared to this variability in the main analysis (assumed to be as observed in pre-control settings). Therefore, we set $k_k$ to 2/3 of the value in the main analysis. Third, because diagnostic performance impacts survey design [4, 16], we evaluated the impact of KK specificity, assuming 99.5% specificity instead of 100% in the main analysis. Finally, to assess
the impact of stricter choices about the maximum allowed probability of making
incorrect decisions, we set the risk of undertreating \( E_{\text{undertreat}} \) to 0.025 instead of
0.05 in the main analysis. As a second step, we also decreased the risk of overtreating
\( E_{\text{overtreat}} \) from 0.25 to 0.1.

Results

Cost-efficient survey design for stopping PC

Fig 1 illustrates the required number of children per school \( n_{\text{children}} \) and the total
survey cost \( C_{\text{tot}} \) as a function of the number of sampled schools \( n_{\text{schools}} \) to reliably
stop PC for the different survey designs for each of the three STH species separately.
The panels on \( n_{\text{children}} \) (Fig 1A, 1C and 1E) highlight three important aspects. First,
fewer children per school are required when more schools are included in the survey.
Second, preparing more smears per sample and collecting more stools per child
further reduces the number of children per school. For example, when making program
decisions for Ascaris infections and sampling 5 schools (Fig 1A), 88 children per
school are required for a \( K K_1 \times 1 \) survey design, while this number reduces to 64, 56,
and 52 when deploying \( K K_1 \times 2, K K_2 \times 1 \) and \( K K_2 \times 2 \) survey designs, respectively. Third,
the minimum required size of the surveys varies across the three STH species. For
instance, when sampling 5 schools and deploying a \( K K_1 \times 1 \) survey design, 88 children
per school are required for Ascaris infections (Fig 1A), while this is 126 for hookworm
infections (Fig 1C) and 80 for Trichuris (Fig 1C).
When focusing on the panels representing the total cost $C_{tot}$ (Fig 1B, 1D and 1F), two additional aspects become apparent. Survey designs involving a second stool ($KK_2 \times 1$ and $KK_2 \times 2$) are more financially demanding, even though fewer children are required to be sampled. Re-taking the aforementioned example (making program decisions for *Ascaris* infections and sampling 5 schools), the total survey costs for $KK_1 \times 1$ survey design equals 2,653 US$ ($n_{children} = 88$), while survey costs are up to 3,781 US$ when deploying a $KK_2 \times 2$ survey design ($n_{children} = 52$). Second, after an initial drop, the total survey costs further increase as a function of number of schools (for $KK_1 \times 1$ and $KK_1 \times 2$ survey designs across all STH species). In other words, there is an optimal number of schools that minimizes the costs while ensuring reliable decision-making.

Combining all these aspects and considering that some schools do not have more than 100 children, $KK_1 \times 2$ is the most cost-efficient of all feasible survey designs for all three STH species, but the difference with $KK_1 \times 1$ is modest. However, the required sample size ($n_{schools} \times n_{children}$) and the total survey costs varies across the 3 STH species (*Ascaris*: 5 x 64, 2,465 US$ (Fig 1B); hookworm: 6 x 66, 2,983 US$ (Fig 1D); *Trichuris*: 4 x 70, 2,022 US$ (Fig 1F)).
Fig 1. The identification of the most cost-efficient survey design to reliably stop preventive chemotherapy. This figure illustrates the required number of children per school ($n_{\text{children}}$; Panels A, C and E) and the corresponding total survey cost ($C_{\text{tot}}$; Panels B, D and F) as a function of the number of sampled schools ($n_{\text{schools}}$) for different survey designs and soil-transmitted helminths (Ascaris: Panels A and B; hookworm: Panels C and D; Trichuris: Panels E and F). The survey designs ($K_{K_a \times b}$) varied in the number of stool samples per child ($= a$) and the number of Kato-Katz thick smears per sample ($= b$). The black bullet point in Panels B, D and F indicates the most cost-efficient survey design. In other words, the number of schools that minimizes the costs while ensuring reliable decision-making. Note that we assumed that the maximum number of children per school could not exceed 100.
In Table 3, we explored the impact of assumptions about inter-individual variation in EPG and the specificity of KK on what is the optimal survey design ($n_{schools} \times n_{children}$) for the different STH species separately. This table indicates three important patterns when considering *Ascaris*. First, compared to an aggregation of infections within children ($k_k$), varying as a function of school mean EPG (5 x 64), fixing $k_k$ (= 1.75) resulted in a smaller sample size (3 x 90) and a reduction of the total survey costs of 11.5% (= 2,181 US$/ 2,465 US$). Second, when assuming a post-control $k_k$ (= 2/3 x varying $k_k$) resulted in slight increase in both sample size (5 x 68) and total survey costs (1.7% = 2,507 US$/ 2,465 US$). Finally, a reduced clinical specificity (= 99.5%) increased the sample size from 5 x 64 to 7 x 62, requiring 38.8% more funds. These patterns were very similar for the other species (Table 3).

We also verified the additional survey costs to further minimize the probability of over- and undertreatment (Table 3). Obviously, further minimizing the risk of undertreatment ($E_{under treat}$) and overtreatment ($E_{overtreat}$) increases the sample size and hence the total survey cost, but not extremely. When we reduce the maximum allowed risk of undertreatment from 0.05 to 0.025, the total survey costs increase by 24% (3,058 US$ vs. 2,465 US$). When we also reduce the risk of undertreatment from 0.25 to 0.10, the sample size further increases to 78.5% (4,400 US$ vs. 2,465 US$). Again, we observed similar patterns for the other two species (Table 3).
Table 3. The impact of alternative model assumptions on both the required sample size and the total survey costs to reliably stop preventive chemotherapy. This table presents the sample size \((n_{\text{schools}} \times n_{\text{children}})\) the decision cut-off \(c\) and the associated total survey costs \((C_{\text{tot}})\) for different scenarios of parameterizing the aggregation of infections within children \((k_k)\), the clinical specificity of Kato-Katz thick smear and the maximum allowed risk of incorrect decision-making for the different soil-transmitted helminths species separately. As a reference we assumed that \(k_k\) varies as a function of school mean eggs per gram of stool (EPG) and that Kato-Katz thick smears has a perfect specificity. We allowed for a risk of undertreatment \((E_{\text{under treat}})\) equal to 0.05 and risk of overtreatment \((E_{\text{overtreat}})\) equal to 0.25. Also, the fixed aggregation parameter (Fixed \(k_k\)), which captures the variation between individuals within schools, was 0.326 for Ascaris, 0.257 for hookworm, and 0.532 for Trichuris. A lower value of \(k_k\) (Lower \(k_k\)) means higher inter-individual variation in EPG values, assumed to be 2/3 of \(k_k\) in the main analysis.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Ascaris</th>
<th>Hookworm</th>
<th>Trichuris</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample size (n_{\text{schools}} \times n_{\text{children}})</td>
<td>(c)</td>
<td>(C_{\text{tot}}) (US$)</td>
</tr>
<tr>
<td>Varying (k_k) as a linear function of school mean EPG</td>
<td>5 x 64</td>
<td>7</td>
<td>2,465</td>
</tr>
<tr>
<td>Fixed (k_k)</td>
<td>3 x 90</td>
<td>7</td>
<td>2,181</td>
</tr>
<tr>
<td>Lower (k_k)</td>
<td>5 x 68</td>
<td>7</td>
<td>2,507</td>
</tr>
<tr>
<td>Parameters</td>
<td>Ascaris</td>
<td>Hookworm</td>
<td>Trichuris</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>Sample size</td>
<td>c</td>
<td>$C_{tot}$ (US$)</td>
</tr>
<tr>
<td>Specificity of Kato-Katz thick smear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>5 x 64</td>
<td>7</td>
<td>2,465</td>
</tr>
<tr>
<td>99.5%</td>
<td>7 x 62</td>
<td>12</td>
<td>3,422</td>
</tr>
<tr>
<td>Risks of incorrect decision-making</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E_{undertreat} = 0.05; E_{overtreat} = 0.25$</td>
<td>5 x 64</td>
<td>7</td>
<td>2,465</td>
</tr>
<tr>
<td>$E_{undertreat} = 0.025; E_{overtreat} = 0.25$</td>
<td>6 x 72</td>
<td>9</td>
<td>3,058</td>
</tr>
<tr>
<td>$E_{undertreat} = 0.025; E_{overtreat} = 0.10$</td>
<td>9 x 62</td>
<td>13</td>
<td>4,400</td>
</tr>
</tbody>
</table>
Cost-efficient survey design to declare elimination of STH as a public health problem

After having assessed study designs and decision cut-offs for stopping PC, we now turn to study designs for declaring elimination of STH as a public health problem (Table 4; Supplement S4, Figure S1). Compared to the most cost-effective survey designs for stopping PC, two important differences can be noted. First, the \( KK_1 \times 1 \) rather than \( KK_1 \times 2 \) survey design is the most cost-efficient choice. Second, the required sample size \((n_{\text{schools}} \times n_{\text{children}})\) is substantially larger (Ascaris: 11 x 84 vs. 5 x 64; hookworm: 25 x 52 vs. 6 x 66; Trichuris: 19 x 68 vs. 4 x 70). Compared to the total survey costs to reliably stop PC, the cost of surveys to reliably declare EPHP are 2.3 (Ascaris) to 4.6 (Trichuris) times higher.

Table 4. The cost-efficient survey design to declare elimination of soil-transmitted helminths as a public health problem. This table represents the required sample size \((n_{\text{schools}} \times n_{\text{children}})\), the decision cut-off \(c\) and the total survey costs \((C_{\text{tot}})\) for four different survey designs \((KK_a \times b)\) that vary in the number of stool samples per child \((= a)\) and the number of Kato-Katz thick smears per sample \((= b)\).

<table>
<thead>
<tr>
<th>STH</th>
<th>Survey design</th>
<th>(n_{\text{schools}} \times n_{\text{children}})</th>
<th>(c)</th>
<th>(C_{\text{tot}}) (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascaris</td>
<td>(KK_1 \times 1)</td>
<td>11 x 84</td>
<td>13</td>
<td>5,775</td>
</tr>
<tr>
<td>Hookworm</td>
<td>(KK_1 \times 1)</td>
<td>25 x 52</td>
<td>17</td>
<td>11,522</td>
</tr>
<tr>
<td>Trichuris</td>
<td>(KK_1 \times 1)</td>
<td>19 x 68</td>
<td>16</td>
<td>9,346</td>
</tr>
</tbody>
</table>
We also explored the impact of the assumptions made on the parameterization of the aggregation in infections between children ($k_k$), the clinical specificity and the allowed risk for both over- and undertreatment on the sample size ($n_{\text{schools}} \times n_{\text{children}}$) and the total survey cost. Not unexpectedly, we observed similar trends as for reliably stopping PC, and hence we refer to Supplement S4, Table S1 for more details.

Discussion

We have demonstrated that the current WHO guidelines to M&E STH control programs may not guarantee reliable decision-making for all program targets and STH species. Based on our results, we rather recommend deploying a duplicate KK on one stool when making decisions to stop PC, deploying a single KK on one stool sample when the aim is to verify whether STH has been eliminated as a public health problem. This difference in survey design is in line with other researchers' findings [10] and can be explained by the fact that examining multiple KK smears increases the sensitivity [5,10,17,18] but generally does not affect estimates of the mean intensity of infection [5,10,17,19]. Also, collecting samples on two consecutive days ($\text{KK}_2 \times 1$ and $\text{KK}_2 \times 2$) was not cost-efficient, because it requires more funds. In addition, as the required sample size varies across STH species (due to differences in fecundity and aggregation of infections; see Table 1), it will be important to ensure sufficient number of schools and children per school to arrive at reliable decision-making for all STH present in the surveyed implementation unit. In other words, the sample size is dictated by the STH species that requires the largest number of schools and children.
This means that in an area where all three STH are present, the sample size to reliably stop PC is dictated by hookworms (6 schools and 66 children per school; see Table 3). Note that STH-specific decision cut-offs vary across sample size, and hence they will need to be adapted accordingly. In Supplement S4, Tables S2 and S3, we therefore provide the required sample size and the corresponding STH-specific decision cut-offs for areas where more than STH species are considered for decision-making.

We showed that the aggregation of infections within children \( (k_k) \) varies as a function of school mean EPG (see Supplement S1), which is in agreement with a recent analysis of the TUMIKIA data. Indeed, this analysis showed that if not considered properly, it leads to poor decision-making due to smaller-than-required sample sizes [20]. Furthermore, the impact of reduced clinical specificity (resulting in increased sample size and total survey cost) was in line with previous work [4], highlighting the need for highly specific diagnostic methods when targeting low-intensity infections settings [21–23].

Minimizing the risk of making incorrect program decisions (risk of overtreatment and undertreatment) is more financially demanding. However, by improving decision-making, one will also save substantial costs due to unnecessarily distributing tablets or re-establishing the PC program. Furthermore, future research using transmission models is needed to assess the appropriateness of the WHO-recommended thresholds for STH policy.
The LQAS approach in this simulation study has also been used for population-based decision-making in schistosomiasis [24] and trachoma [25] control programs. Conversely, another approach is the model-based geostatistical methods for M&E of STH, and other NTDs [26,27]. Yet, compared to the LQAS method, this alternative approach adoption required additional expertise in geostatistical analysis [4,27].

Conclusion

We confirm that KK is valuable for M&E of STH in low-endemicity infections settings, but the current WHO-recommended survey design may not guarantee reliable decision-making to stop PC, and especially not for declaring EPHP. We, therefore, provide cost-efficient alternative survey designs.
References


7. Niguse AF, Hailu T, Alemu M, Nibret E, Amor A, Munshea A. Evaluating the


Funding

Bruno Levecke acknowledges funding from Ghent University starting grant (www.ugent.be) and Luc E. Coffeng from the Dutch Research Council (NWO, grant 016.Veni.178.023). In addition, the financial support for the dataset used in this study was provided by Schistosomiasis Control Initiative, the Partnership for Child Development, the Children's Investment Fund Foundation, the End Neglected Tropical Diseases Fund, and UKAID-DFID.
537 Supporting Information Legends

540 **Supplementary information S1.** Quantification of between and within-school variation in faecal egg counts

543 **Supplementary information S2.** The mathematical backbone of the simulation framework

546 **Supplementary information S3.** Estimation of total survey cost to monitor and evaluate STH when using Kato-Katz thick smear

549 **Supplementary information S4.** Additional tables and figures