

1 A general framework to support cost-efficient fecal egg
2 count method and study design choices for large-scale STH
3 deworming programs – monitoring of therapeutic drug
4 efficacy as a case study

5

6 **Running title:** Towards evidence-based recommendations for STH programs

7

8 Luc E Coffeng^{1,#,*}, Johnny Vlaminck^{2,#}, Piet Cools³, Matthew Denwood⁴, Marco Albonico⁵, Shaali M
9 Ame⁶, Mio Ayana^{3,7}, Daniel Dana^{3,7}, Giuseppe Cringoli⁸, Sake J de Vlas¹, Alan Fenwick⁹, Michael
10 French^{9,10}, Adama Kazienga², Jennifer Keiser^{11,12}, Stefanie Knopp^{11,12}, Gemechu Leta¹³, Leonardo
11 F. Matoso^{14,15}, Maria P. Maurelli⁸, Antonio Montresor¹⁶, Greg Mirams¹⁷, Zeleke Mekonnen⁷,
12 Rodrigo Corrêa-Oliveira¹⁴, Simone A. Pinto¹⁴, Laura Rinaldi⁸, Somphou Sayasone¹⁸, Peter
13 Steinmann^{11,12}, Eurion Thomas¹⁹, Jozef Vercruyse² and Bruno Levecke^{2,*}

14

15 ¹ Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam,
16 The Netherlands

17 ² Department of Translational Physiology, Infectiology and Public Health, Ghent University,
18 Merelbeke, Belgium

19 ³ Department of Diagnostic Sciences, Faculty of Medicine and Health Sciences, Ghent University,
20 Ghent, Belgium

21 ⁴ Department of Veterinary and Animal Sciences, University of Copenhagen, Denmark

22 ⁵ National Health System, Turin, Italy

23 ⁶ Laboratory Division, Public Health Laboratory-Ivo de Carneri, Chake Chake, United Republic of
24 Tanzania

25 ⁷ Jimma University Institute of Health, Jimma University, Jimma, Ethiopia

26 ⁸ Department of Veterinary Medicine and Animal Production, University of Naples Federico II,
27 Naples, Italy

28 ⁹ Schistosomiasis Control Initiative, Department of Infectious Disease Epidemiology, St Mary's
29 Campus, Imperial College London, London, United Kingdom

30 ¹⁰ RTI International, Washington DC, USA

31 ¹¹ Swiss Tropical and Public Health Institute, Allschwil, Switzerland

32 ¹² University of Basel, Basel, Switzerland

33 ¹³ Ethiopian Public Health Institute, Addis Ababa, Ethiopia

34 ¹⁴ Laboratory of Molecular and Cellular Immunology, Research Center René Rachou - FIOCRUZ,
35 Belo Horizonte, Brazil

36 ¹⁵ Nursing school, Federal University of Minas Gerais, Brazil

37 ¹⁶ Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva,
38 Switzerland

39 ¹⁷ Techion Group Ltd, Dunedin, New Zealand

40 ¹⁸ Lao Tropical and Public Health Institute, Ministry of Health, Vientiane, Lao People's Democratic

41 Republic

42 ¹⁹ Techion Group Ltd, Aberystwyth, United Kingdom

43

44 *Corresponding authors: l.coffeng@erasmusmc.nl (LEC) and bruno.levecke@ugent.be (BL)

45

46 #These authors contributed equally to this work.

47 Abstract

48 **Background:** Soil-transmitted helminth (STH) control programs currently lack evidence-based
49 recommendations for cost-efficient survey designs for monitoring and evaluation. Here, we
50 present a framework to provide evidence-based recommendations, using a case study of
51 therapeutic drug efficacy monitoring based on the examination of helminth eggs in stool.

52 **Methods:** We performed an in-depth analysis of the operational costs to process one stool sample
53 for three diagnostic methods (Kato-Katz, Mini-FLOTAC and FECPAK^{G2}). Next, we performed
54 simulations to determine the probability of detecting a truly reduced therapeutic efficacy for
55 different scenarios of STH species (*Ascaris lumbricoides*, *Trichuris trichiura* and hookworms), pre-
56 treatment infection levels, survey design (screen and select (*SS*); screen, select and retest (*SSR*)
57 and no selection (*NS*)) and number of subjects enrolled (10 – 2,000). Finally, we integrated the
58 outcome of the cost assessment into the simulation study to estimate the total survey costs and
59 determined the most cost-efficient survey design.

60 **Principal Findings:** Kato-Katz allowed for both the highest sample throughput and the lowest cost
61 per test, while FECPAK^{G2} required both the most laboratory time and was the most expensive.
62 Counting of eggs accounted for 23% (FECPAK^{G2}) or ≥80% (Kato-Katz and Mini-FLOTAC) of the total
63 time-to-result. *NS* survey designs in combination with Kato-Katz were the most cost-efficient to
64 assess therapeutic drug efficacy in all scenarios of STH species and endemicity.

65 **Conclusions/significance:** We confirm that Kato-Katz is the fecal egg counting method of choice
66 for monitoring therapeutic drug efficacy, but that the survey design currently recommended by
67 WHO (*SS*) should be updated. Our generic framework, which captures laboratory time and
68 material costs, can be used to further support cost-efficient choices for other important surveys

69 informing STH control programs. In addition, it can be used to explore the value of alternative
70 diagnostic techniques, like automated egg counting, which may further reduce operational costs.

71 **Author Summary**

72 Large-scale deworming programs are implemented worldwide to reduce morbidity caused by
73 intestinal worms. As these programs operate in resource-poor-settings, it is important that their
74 operational costs are minimized without jeopardizing the quality of decision-making. We present
75 a framework for evidence-based recommendations for cost-efficient decision-making in
76 deworming programs, using monitoring of therapeutic drug efficacy via stool examination as a
77 case study. To this end, we first assessed the time and the cost of processing stool samples in a
78 laboratory according to different diagnostic methods. Then for each diagnostic method, survey
79 design, and a range of settings (pre-dominant worm species and pre-treatment infection levels),
80 we calculated the probability of correctly detecting a truly reduced therapeutic drug efficacy and
81 the associated operational costs. Generally, the estimated operational costs varied across
82 diagnostic method, survey design, worm species and disease endemicity. Based on our findings,
83 we conclude that the use of the current diagnostic standard, the Kato-Katz method, is justified to
84 assess drug efficacy, but that a change in survey design is warranted. Our methodology, which
85 leverages detailed data on laboratory time and material costs, can also be used to provide
86 evidence-based recommendations for other types of decisions in large-scale deworming
87 programs.

88

89 **Key words:** soil-transmitted helminths, Kato-Katz thick smear, Mini-FLOTAC, FECPAK^{G2}, cost
90 analysis, preventive chemotherapy, anthelmintic resistance, egg reduction rate, survey designs

91 Introduction

92 Soil-transmitted helminths (STHs; *Ascaris lumbricoides*, *Trichuris trichiura* and the hookworm
93 species *Necator americanus* and *Ancylostoma duodenale*) infect approximately 800 million
94 individuals across the world and are responsible for the loss of more than three million disability-
95 adjusted life years annually [1,2]. To control morbidity associated with these infections, the World
96 Health Organization (WHO) strives to reduce the prevalence of moderate-to-heavy intensity (MHI)
97 infections to less than 2% [3]. To reach this goal, anthelmintic drugs are periodically distributed to
98 at-risk populations through large-scale deworming programs – so-called preventive chemotherapy
99 [4]. In these programs, periodic follow-up surveys are conducted to determine whether the
100 therapeutic efficacy of the administered drugs is still satisfactory [5], and whether stopping or
101 scaling down drug administration is justified [6]. However, as these programs often operate in
102 resource-poor settings, it is important to minimize operational costs without jeopardizing the
103 correctness of the program decisions (e.g., avoiding prematurely scaling down of preventive
104 chemotherapy or continuing the administration of anthelmintic drugs with a reduced therapeutic
105 drug efficacy). An important proportion of STH survey costs is related the processing of stool
106 samples and the counting of STH eggs under a microscope. Speich and colleagues [7]
107 demonstrated that, independent of the evaluated diagnostic method, the lion's share (~70%) of
108 the total costs of performing egg counts in Zanzibar was made up of salaries. More recently, Leta
109 and colleagues calculated those personnel salaries (~40%) and car rental fees (~50%) made up a

110 combined ~90% of the total study costs when doing a national STH mapping survey in Ethiopia [8].
111 Hence, the number of samples to be screened, the speed at which technicians can process a single
112 sample, the number of samples that can be processed per day, and thus the number of sampling
113 days, are considered the major cost drivers of programmatic surveys for infection prevalence or
114 therapeutic drug efficacy.

115
116 Several different microscopy-based methods (e.g., Kato-Katz thick smear (KK), Mini-FLOTAC,
117 McMaster and FECPAK^{G2}) are used to diagnose STH infections in stool, of which some are more
118 complex than others [9-11]. Of all currently applied methods, the WHO-endorsed KK is the most
119 widely established. This method produces smears of 41.7 mg of stool to visualize STH eggs for
120 microscopic identification and counting [12] and is thought to be relatively easy and affordable
121 [7]. The Mini-FLOTAC employs a flotation solution to separate STH eggs from stool debris in a
122 special device prior to counting [9]. The FECPAK^{G2} method is the most recent and innovative
123 diagnostic method [11, 13, 14]. It is also a flotation-based method, but instead of using a standard
124 microscope, it employs a purpose-made device to accumulate STH eggs in one field of view, and
125 to produce a digital image of this view that can later be marked up by a technician [11]. However,
126 in a previous study, it was shown that both FECPAK^{G2} and Mini-FLOTAC had a clinical sensitivity
127 equal or inferior to a single KK for all STHs [15], and that these flotation-based methods provided
128 lower fecal egg counts (FECs; expressed as eggs per gram of stool (EPG)) compared to KK [16].

129
130 Making an evidence-based choice about which FEC method to use in STH control programs
131 remains non-trivial, particularly when a decision-making framework is intended to be applied to a

132 wide range of epidemiological settings. This is because the suitability and the cost of different
133 survey designs and diagnostic techniques will vary by epidemiological setting [17]. For instance,
134 the probability of making correct policy decisions may strongly depend on both the performance
135 of a particular diagnostic method and the associated decision criterion in a particular
136 epidemiological setting [18-23]. For STH, this performance depends on the average intensity of
137 infection in a community as well as the level of variation in egg excretion (between individuals and
138 within individuals over time), and in the case of the evaluation of drug efficacy, variation in
139 individual drug responses [20,23]. Further, it is important to consider that the total operational
140 cost of a survey will depend on the consumable costs of the diagnostic method used, the survey
141 design (number of samples and number of days spent in the field), and the time needed to count
142 eggs [20]. Importantly, the latter will depend on how many eggs need to be counted, which has
143 not been considered before and which will vary by epidemiological setting and will depend on the
144 goal of the survey (e.g., detecting infection (counting at least one egg) or quantifying intensity of
145 infection (counting all eggs)). Quantifying these costs requires an in-depth analysis of the
146 operational costs of processing samples with different FEC methods.

147
148 We aim to provide a general framework for evidence-based recommendations for cost-efficient
149 decision-making in large-scale STH deworming programs based on FEC methods, using monitoring
150 of therapeutic drug efficacy as a case study. To this end, we performed an in-depth analysis of the
151 operational costs to process one sample for three FEC methods (KK, Mini-FLOTAC and FECPAK^{G2})
152 based on the time-to-result and an itemized cost assessment. Next, we performed a simulation
153 study to determine the probability of correctly concluding that the therapeutic drug efficacy is

154 reduced based on different FEC methods, survey designs and numbers of individuals enrolled,
155 while accounting for the variation in both egg counts and individual drug responses. Finally, we
156 integrated the outcome of the in-depth cost-assessment into the simulation study to determine
157 the most cost-efficient diagnostic test and survey design to detect presence of reduced drug
158 efficacy for the different STH species across different scenarios of STH endemicity.

159 **Methods**

160 **Ethics statement**

161 Data were collected from four sites during a drug efficacy trial designed to test the equivalence of
162 different FEC methods in attaining estimates of the therapeutic efficacy of a single oral dose of
163 400 mg albendazole (ALB) against STH infections in school aged children (SAC) [24]. The trial was
164 performed in Brazil, Ethiopia, Lao PDR and Zanzibar (Pemba Island). The study protocol for this
165 trial were reviewed and approved by the Ethics Committee of the Faculty of Medicine and Health
166 Sciences, the University Hospital of Ghent University, Belgium (Ref. No B670201627755;
167 2016/0266) and the national ethical committees associated with each trial site (Ethical Review
168 Board of Jimma University, Jimma, Ethiopia: RPGC/547/2016; National Ethics Committee for
169 Health Research (NECHR), Vientiane, Lao PDR: 018/NECHR; Zanzibar Medical Research and Ethics
170 Committee, United Republic of Tanzania: ZAMREC/0002/February/2015; and the Institutional
171 Review Board from Centro de Pesquisas René Rachou, Belo Horizonte, Brazil: 2.037.205). The trial
172 was retrospectively registered on Clinicaltrials.gov (ID: NCT03465488) on March 7, 2018.

173 In-depth analysis of the operational costs to process one sample for three FEC
174 methods based on the time-to-result and an itemized cost assessment

175 **Time-to-result**

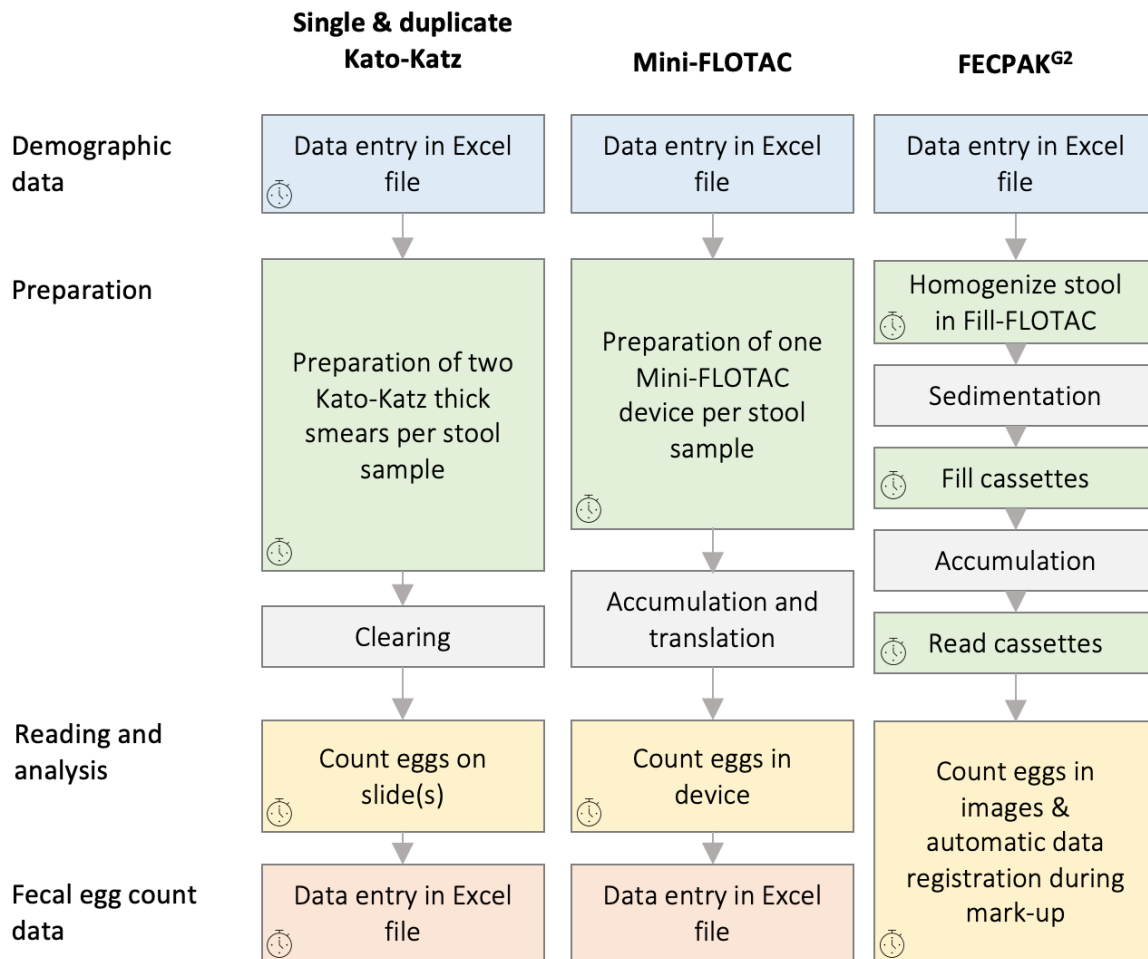
176 Measuring time-to-result for the different FEC methods was part of the drug efficacy trial, which
177 have been extensively described elsewhere [24,25]. During the trial, baseline stool samples were
178 collected from SAC, who were subsequently treated with a single dose of 400 mg ALB. Between
179 14 and 21 days after treatment, SAC who were positive for any STH species at baseline were re-
180 sampled to evaluate the reduction in egg output (ERR). At baseline and follow-up, stool samples
181 were processed by duplicate KK (slide A and B), Mini-FLOTAC and FECPAK^{G2} to determine FECs
182 (expressed in EPG) for each STH separately.

183

184 Upon arrival in the laboratory, stool samples were first grouped into batches of ten samples (with
185 the remainder in a separate last batch). Subsequently, all stool samples were homogenized by
186 stirring with a wooden tongue depressor. Finally, subsamples were taken to be processed
187 according to the different FEC methods. **Fig 1** provides an overview of the different steps timed
188 for each FEC method, including preparing the sample for analysis, counting eggs and data entry
189 (demographic data and FECs), which ultimately resulted in the time-to-result measurement.

190 **Fig 1. Overview of the different operational steps for the different FEC methods.** The distinctive steps to perform a
 191 Kato-Katz (KK), Mini-FLOTAC or FECPAK^{G2} on a single stool sample are provided in chronological order per method.
 192 The procedures are grouped per main subject (**blue**: entry of demographic data; **green**: preparation of the sample;
 193 **yellow**: reading of the slide/device or the image to count STH eggs; **red**: entry of fecal egg count data). Waiting steps
 194 included in the procedure are indicated in **grey** and represent a fixed amount of time. The small clock symbol indicates
 195 what steps have been timed as part of this experiment.

196



197

198

199 Detailed standard operating procedures (SOPs) to time the preparatory steps and the egg counting
 200 process are described elsewhere (see S3-5 info of Vlaminck *et al.* [24]); **S1 Info** provides a brief
 201 summary). A summary of the SOP to time the data entry is provided in **S2 Info**.

202 We expressed the time (in seconds) needed to enter data and prepare samples for analysis per

203 batch by dividing the total time recorded per batch by the number of samples within that batch.

204 These calculations included batches gathered at baseline and follow-up. Batches containing fewer
205 than 5 samples were not timed and were excluded from these calculations. We report the average
206 reading time, preparation time and data entry time across batches. The overall mean preparation
207 time per sample was calculated as the mean of batch-specific estimates of time per sample. The
208 data on the timing of egg counting were analyzed at the level of samples. For each FEC method,
209 the correlation between the time required to count and the absolute number of eggs in the sample
210 was quantified using linear regression models. In these models, we predicted the \log_{10} -
211 transformed time (in seconds) needed for egg counting (dependent variable) using the square of
212 the \log_{10} -transformed total number of STH eggs counted plus 1 as the independent variable.
213 Statistical analyses were conducted in R [26], Microsoft Excel v16.16 .7 and Prism version 6.0. for
214 Mac.

215

216 **Itemized cost assessment**

217 We calculated the cost of materials to collect a stool sample in a school setting and the costs to
218 perform the FEC method, including the required equipment, supplies and reagents. For this, we
219 performed an itemized cost assessment considering the cost per unit, the usage over a one-year
220 period and the expected duration of use (in years). A detailed itemized cost assessment to collect
221 stool samples and to perform the FEC method is provided in **S3 Info**. For specific items, such as
222 the KK kit, Mini-FLOTAC or FECPAK^{G2} devices, we used the prices that were either advertised online
223 or obtained through the manufacturer (2020). To estimate the cost of everyday materials, such as
224 scissors, paper, salt and buckets, Ethiopian market prices were used (2020).

225

226 Simulation study to assess the probability of correctly detecting a truly reduced
227 therapeutic efficacy

228 **Definition of a reduced therapeutic efficacy and survey designs**

229 For each candidate survey design, we determined the probability that the resulting ERR point-
230 estimate confirmed the presence of reduced efficacy of a single oral dose of 400 mg ALB. Here,
231 we assumed that the true efficacy was 5%-points under the species-specific thresholds specified
232 by WHO (**Table 1**) [5], and we concluded the presence of reduced therapeutic efficacy if the ERR
233 point-estimate was under the WHO threshold. Given that the endemicity at baseline has an impact
234 on the statistical power [27] and the total survey costs [21], we determined the probability of
235 correctly detecting a truly reduced therapeutic efficacy across different scenarios of endemicity.

236
237 Currently, it is recommended by WHO to determine the efficacy based on individuals that were
238 egg-positive at baseline only; however, excluding individuals who were egg-negative at baseline
239 from the analysis may result in a substantial overestimation of drug efficacy due to regression to
240 the mean, particularly in low endemic settings or when true drug efficacy is low. Coffeng and
241 colleagues showed previously that this bias could be avoided by a number of alternative survey
242 designs [21]. However, an in-depth analysis of the associated operational costs was missing. Here,
243 we re-evaluate some of the survey designs assessed by Coffeng and colleagues, including the
244 WHO-recommended ‘screen and select’ design (*SS*; only egg-positive individuals are followed up),
245 the ‘screen, select, and retest’ (*SSR*; only egg-positive individuals are followed up, but the analysis
246 is based on a second separate baseline stool sample [21]), and the ‘no selection’ design (*NS*; all
247 enrolled individuals are screened at baseline and follow-up). For the *NS* and *SSR* survey design,

248 we explored two variants, one that was based on a single FEC on the follow-up sample ($NS_{1 \times 1 / 1 \times 1}$
249 and $SSR_{1 \times 1 / 1 \times 1}$), and one that was based on duplicate FECs ($NS_{1 \times 1 / 1 \times 2}$ and $SSR_{1 \times 1 / 1 \times 2}$). We did
250 not consider survey designs based on a single FEC on two consecutive stool samples at follow-up
251 FEC ($NS_{1 \times 1 / 2 \times 1}$ and $SS_{1 \times 1 / 2 \times 1}$), as sample collection on two consecutive days adds considerable
252 logistical issues while yielding relatively little in terms of precision in drug efficacy estimates [21].
253 For the WHO-recommended SS design we considered two variants: one based on a single FEC
254 both at baseline and follow-up ($SS_{1 \times 1 / 1 \times 1}$), and one based on duplicate FECs at both time points
255 ($SS_{1 \times 2 / 1 \times 2}$). The former is recommended in the WHO manual to monitor the therapeutic efficacy
256 of drugs against schistosomes and STH [5], and the latter is currently being piloted in a number of
257 endemic countries as part of the Starworms project [27].

258

259 **General simulation framework**

260 For the current simulation study, we adapted the framework described by Coffeng and colleagues
261 [21], accounting for the following sources of variation in egg counts:

- 262 1. Inter-individual variability in mean egg intensity due to variation in infection levels
263 between individuals (assumed to follow a gamma distribution);
- 264 2. Day-to-day variability in mean egg intensity within an individual due to heterogeneous
265 egg excretion over time (assumed to follow a gamma distribution);
- 266 3. Variability in egg counts between repeated aliquots of a stool sample due to the
267 aggregated distribution of eggs in stool (assumed to follow a Poisson or a gamma-Poisson
268 (i.e., negative binomial) distribution);

269 4. Inter-individual variability in the effect of drug administration in terms of the ERR
 270 (assumed to follow a beta distribution).

271 For the quantification of each gamma distribution, we followed the approach of Denwood et al
 272 [28] in using the coefficient of variation (cv) as a standardised measure of variability, which is
 273 related to the shape parameter k of a gamma distribution by taking $k = cv^{-2}$. Species-specific
 274 variability between individuals and within individuals over time were estimated based on data
 275 from clinical trials during which a duplicate KK was performed on two consecutive stool samples
 276 both at baseline and follow-up [29]. STH species and FEC method-specific variability between
 277 repeated aliquots of the same stool sample were estimated from the egg count data published by
 278 Cools et al. [30]. The average difference between FEC methods in terms of egg recovery
 279 performance was also determined as flotation techniques are known to miss unfertilized eggs [31]
 280 (see **S4 Info** for details). The parameterization of the simulation framework is summarized in **Table**
 281 **1**. For a detailed description of the simulation model we refer to **S5 Info**.

282

283 **Table 1. Parameterization of the simulation framework for variability in fecal egg counts before and after treatment.**

Parameter	<i>Ascaris lumbricoides</i>	<i>Trichuris trichiura</i>	Hookworm
Mean EPG in the population (μ) as measured by KK			
Endemicity level 1: 1.0 - 9.9%	9.6	2.8	3.7
Endemicity level 2: 10.0 - 19.9%	85.2	12.9	23.7
Endemicity level: 20.0 - 49.9%	360.0	49.7	61.7
Endemicity level: 50.0 - 100.0%	2195.5	124.7	210.3
Variability in EPG between individuals (shape k_i)	0.327	0.444	0.250
Day-to-day variability in EPG within individuals (shape k_d)	0.510	1.000	1.000
Weight of stool aliquot (w_s ; gram)			
KK	1/24	1/24	1/24
Mini-FLOTAC	1/10	1/10	1/10
FECPAK ^{G2}	1/34	1/34	1/34
Relative recovery of eggs from stool (ρ_s)			

Parameter	<i>Ascaris lumbricoides</i>	<i>Trichuris trichiura</i>	Hookworm
KK (reference)	1.000	1.000	1.000
Mini-FLOTAC	0.645	1.005	0.801
FECPAK ^{G2}	0.248	0.152	0.569
Variability in egg counts between aliquots based on the same stool sample (shape k_s)			
KK	∞^*	∞^*	∞^*
Mini-FLOTAC	0.579	3.022	1.465
FECPAK ^{G2}	0.520	0.706	0.574
Assumed true mean and variability (95%-CI) in drug efficacy (beta distribution with shape parameters α_{ERR} and β_{ERR})	0.80 (0.69–0.89)	0.35 (0.25–0.45)	0.75 (0.64–0.84)
α_{ERR}	49.5	30.2	53.8
β_{ERR}	12.4	56.1	17.9
WHO definition for reduced therapeutic efficacy based on ERR [5]	<85%	<40%	<80%

284 * Implies $cv_{KK} = 0$, which means that KK-based egg counts from repeated aliquots of the same homogenized stool
285 sample follow a Poisson distribution (no extra over-dispersion in contrast to Mini-FLOTAC or FECPAK^{G2}).
286

287 Using this framework, we simulated egg counts for all survey designs across four scenarios, each
288 representing different population average baseline FECs. The selection of these scenarios was
289 based on infection levels in the nationwide mapping of STH infections in Ethiopia [32] (see also **Fig**
290 **S1**), where each scenario represented the median of school-level mean FEC (in EPG) of one of four
291 endemicity levels (prevalence between 1.0 and 9.9%, 10.0 and 19.9%, 20.0% and 49.9%, or
292 $\geq 50.0\%$). For each survey design, we considered a range of 100 to 5,000 individuals (with
293 increments of 5 individuals) that are initially tested at baseline. For each survey design, sample
294 size and endemicity scenario, 10,000 repeated Monte Carlo simulations were performed. In each
295 simulation, the group-based arithmetic mean ERR was calculated using the recommended
296 procedure [33], and the ERR was considered reduced if under the STH species-specific threshold
297 (**Table 1**). For each survey design and sample size, we then calculated the proportion of the 10,000
298 repeated Monte Carlo simulations that correctly identified therapeutic efficacy as truly reduced
299 ($prob_{reduced}$). If a baseline survey resulted in fewer than 50 egg-positive individuals, the survey

300 was considered to have failed and was discontinued. In those cases, it was considered to not have
301 detected reduced efficacy. In the remainder of the text, the proportion of surveys that fail will be
302 referred to as the “failure rate”. All simulations and calculations were performed using the eggSim
303 package [19] in R [26]. This package allows the same calculations to be made for any arbitrary set
304 of parameter values using highly performant C++ code, and is freely available [34].

305

306 Total operational costs to monitor drug efficacy

307 For each simulated survey, we calculated the total operational costs in terms of (i) the cost of
308 consumables to collect and process samples, (ii) the salary for a single mobile field team comprised
309 of one nurse and three laboratory technicians, and (iii) the cost of transport, including car rental,
310 salary of the driver, and gasoline. We assumed that a working day consists of 8 working hours, and
311 that the daily salary of one team was 80 US\$ (4 per diems of 22.5 US\$) and that the daily cost for
312 transport was 90 US\$. Second, we assumed that the team collected samples in the morning (8:00
313 – 12:00), and that all collected samples were processed in the afternoon (13:00 – 17:00). Complete
314 analysis of all samples on the same day implies that the number of samples that can be collected
315 daily is limited, and that this number will vary across FEC methods, phase of the trial (less time for
316 egg counting is required in follow-up samples) and endemicity (more time for egg counting is
317 required in highly endemic areas). Note that we do not consider costs for the establishment and
318 maintenance of laboratory infrastructure. We further assumed that all work takes place on regular
319 working days, that the team does not take any breaks during processing, and that all samples are
320 collected from a single school/community without loss to follow-up. All cost calculations were

321 based on the itemized cost-assessment described above. Technical details on how the total costs
322 were calculated can be found in **S6 Info**.

323

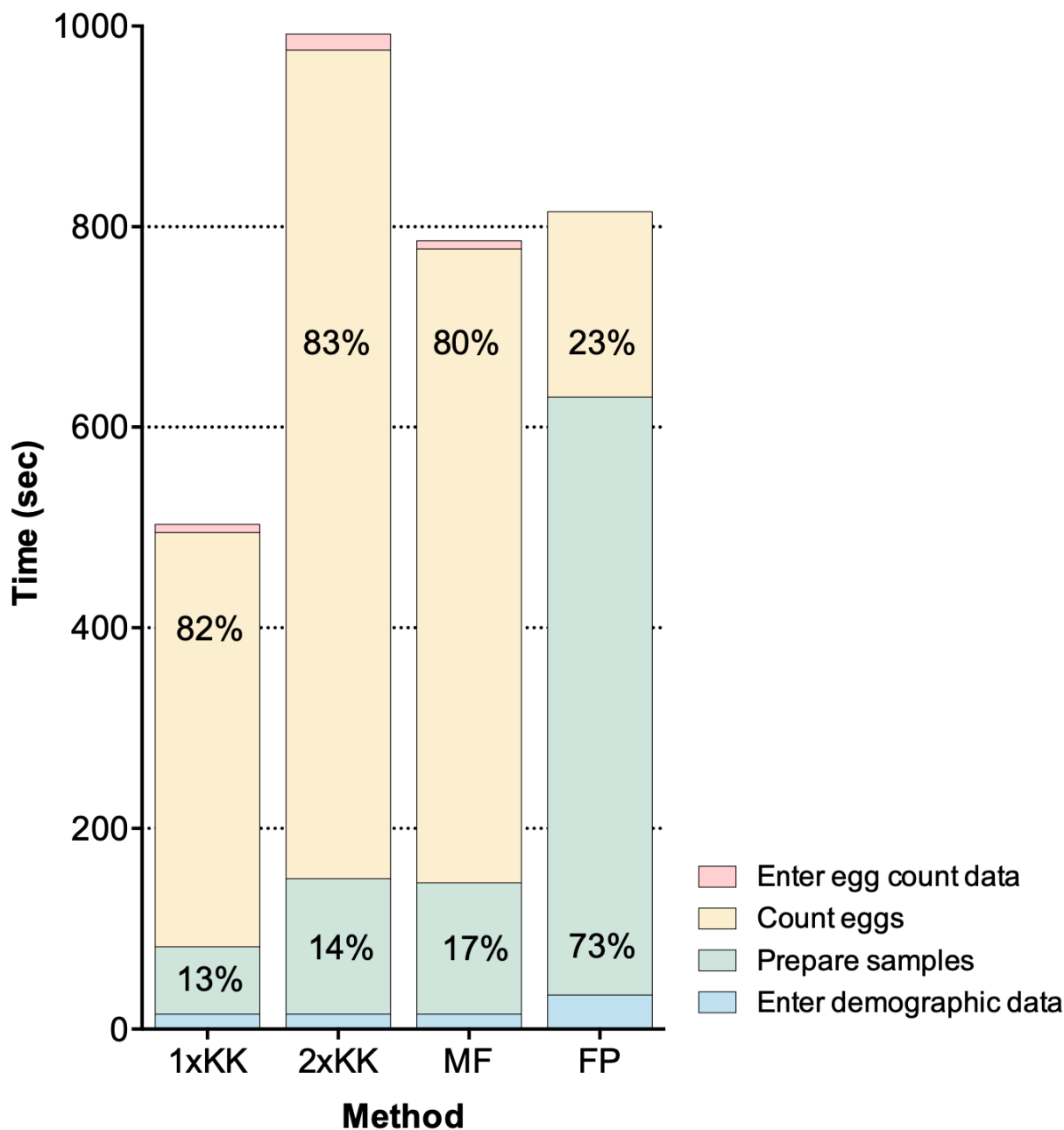
324 **Results**

325 **Time-to-result**

326 During the drug efficacy trial surveys performed in Brazil, Ethiopia, Lao PDR, and Tanzania, we
327 assessed the mean time-to-result (i) to prepare stool samples ($T_{prep,X}$, X representing the number
328 of aliquots per sample), (ii) to count eggs, (iii) to digitize demographic data ($T_{demography}$), and (iv)
329 to digitize the FEC results ($T_{record,X}$). The time analysis is illustrated in **Fig 2**. Overall, a duplicate KK
330 consumed the most time, requiring on average (standard deviation) 989 sec (449). A single KK
331 consumed the least amount of time and required on average 507 sec (318). The mean time-to-
332 result for a single Mini-FLOTAC or FECPAK^{G2} method were 786 sec (513) and 802 sec (329),
333 respectively. The percentage of time-to-result spent on the egg counting process was
334 approximately 80% for both KK (single KK: 413 sec out of 507sec; duplicate KK: 820 sec out of 989
335 sec) and Mini-FLOTAC (632 sec out of 786 sec), while this was 23% for FECPAK^{G2} (185 sec out of
336 802 sec). For the latter method, most of the time-to-result (74%) was spent preparing the samples
337 for analysis (596 sec out of 802 sec).

338 **Fig 2: Time required to quantify soil-transmitted helminth infections in stool by four fecal egg count methods.** The
339 height of the bars represents the mean time (in sec) needed to enter demographic data (**blue**), to perform the
340 preparation phase (**green**), to count eggs (**yellow**) and to enter egg count data (**red**) for a single (1xKK) and duplicate
341 Kao-Katz (2xKK), Mini-FLOTAC (MF) and FECPAK^{G2} (FP). The relative proportion (in %) of total time required to perform
342 the preparation phase and to count is reported inside the bars.

343



344

345

346 As expected, counting a larger number of STH eggs required more reading time, where the log

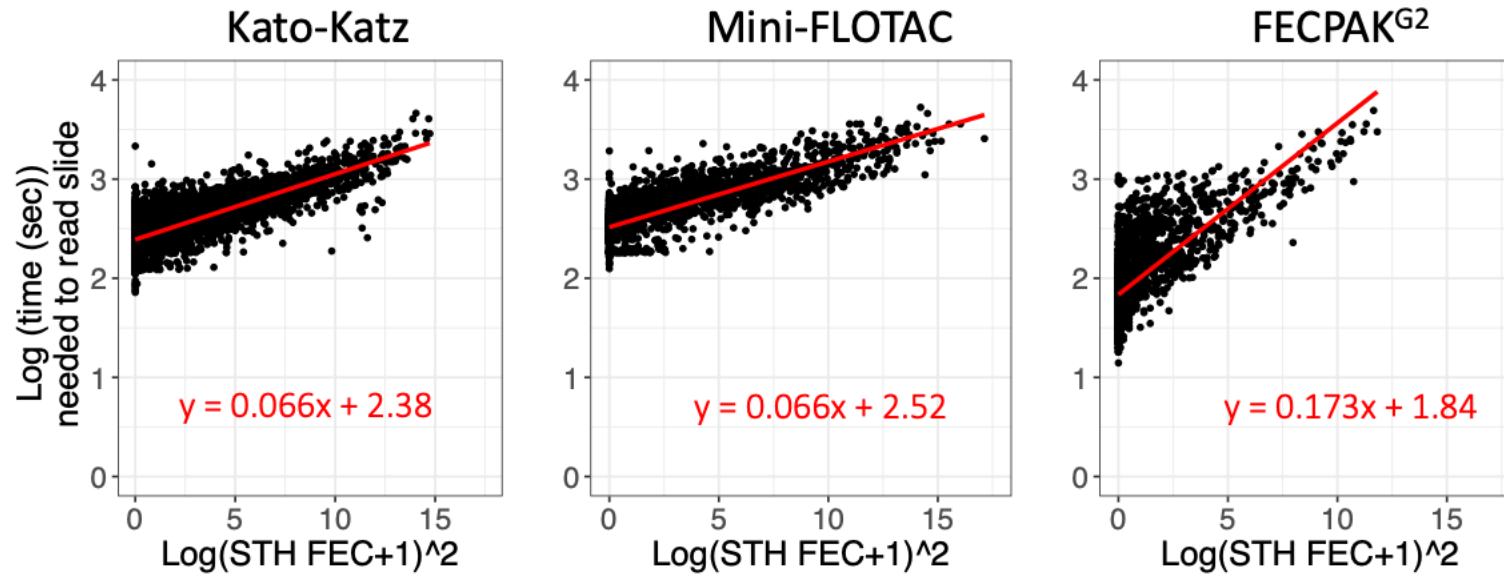
347 transformed total time required to count all eggs could be well described as a linear function of

348 the square of the base-10 log-transformed total egg counts (Fig 3). Table 2 summarizes the

349 average time required for the different steps included in the total survey costs for each FEC

350 method separately. To obtain an estimate for a single KK preparation we divided the average time
351 to prepare a duplicate KK (135 sec) by two (= 67 sec). As a second Mini-FLOTAC can be filled from
352 the same Fill-FLOTAC (no need to weigh and homogenize the sample for a duplicate Mini-FLOTAC),
353 we multiplied the mean time required to process a single Mini-FLOTAC (131 sec) by 1.5 to estimate
354 the time for a duplicate Mini-FLOTAC (197 sec). To estimate the time for a duplicate FECPAK^{G2}, we
355 doubled the time for each of the different preparatory steps, except for the step to prepare the
356 samples in the FECPAK^{G2} sedimenters, resulting in a total time of 1,050 sec (= 142 sec + 2 x 174
357 sec + 2 x 280 sec). Similarly, the mean time needed to enter one duplicate KK result was 18 sec;
358 the time needed to record FEC results based on single KK and Mini-FLOTAC was assumed to be
359 half that value (9 sec). For the FECPAK^{G2} method, no FEC data entry was required as the software
360 automatically registers and stores mark-up data. **S7 Info** provides more detailed information
361 (number of batches timed; the average units per batch; average and SD time) on each step of the
362 sample analysis process, starting with the timing of the demographic data entry followed by the
363 preparation phase, the egg counting process, and the time it took to enter FEC data.

364 **Fig 3. The reading time as a function of the number of STH eggs counted in a sample.** This figure represents the reading time as a function of the number of STH
365 eggs counted in a sample for single Kato-Katz (KK), Mini-FLOTAC and FECPAK^{G2} separately. All egg counts represent raw egg counts (not in eggs per gram of stool).
366 The red line represents the linear regression line. The function of the regression line is also provided.
367



368
369

370 **Table 2. Overview of parameters that determine the time required to process a single stool sample.**

Symbol	Description	Average time required (seconds)		
		KK	Mini-FLOTAC	FECPAK ^{G2}
$T_{demography}$	Time to enter demographic data	15	15	34
$T_{prep,X}$	Time to prepare a stool sample			
$X = 1$	One aliquot	67	131	596
$X = 2$	Two aliquots	135	197	1,050
$f(c)$	Time required to count c eggs (not EPG) in a single aliquot	$10^{(2.3896 + 0.0661 \times \log_{10}(c + 1)^2)}$	$10^{(2.5154 + 0.0661 \times \log_{10}(c + 1)^2)}$	$10^{(1.8349 + 0.1731 \times \log_{10}(c + 1)^2)}$
$T_{record,X}$	Time to record count data			
$X = 1$	One aliquot	9	9	0
$X = 2$	Two aliquots	18	18	0

371

372 **Itemized cost assessment**

373 The costs associated with the materials for stool sample collection in schools and to process stool

374 samples for a single or duplicate KK, mini-FLOTAC and FECPAK^{G2} are reported in detail in **Table 3**.

375 In summary, the costs associated with sampling a single sample ($cost_{sample}$) was US\$ 0.57. The

376 material costs to perform a single FEC ($cost_{aliquot,1}$) were US\$ 1.37 for KK, US\$ 1.51 for the mini-

377 FLOTAC method, and US\$ 1.69 for the FECPACK^{G2} method. When a duplicate FEC was performed

378 on the same sample the material costs ($cost_{aliquot,2}$) were US\$ 1.51 (KK), US\$ 1.87 (Mini-FLOTAC),

379 and US\$ 2.73 (FECPAK^{G2}).

380

381 **Table 3. Overview of cost per unit of consumables, salary and travel.**

Symbol	Description	Cost (US\$)		
		KK	Mini-FLOTAC	FECPAK ^{G2}
$cost_{sample}$	Cost of collecting a single stool sample	0.57	0.57	0.57
$cost_{aliquot,X}$	Cost per aliquot when preparing X aliquots from the same stool sample			
$X = 1$	One aliquot	1.37	1.51	1.69
$X = 2$	Two aliquots	1.51	1.87	2.73
$salary_{perdiem}$	Daily salary for every technician and nurse on the mobile Team	22.50	22.50	22.50

Symbol	Description	Cost (US\$)		
		KK	Mini-FLOTAC	FECPAK ^{G2}
$travel_{perdiem}$	Daily cost of car rental, including petrol and driver wages	90.00	90.00	90.00

382

383 Failure rate, probability of correctly detecting reduced therapeutic efficacy and the
 384 corresponding total survey costs

385 Given the large number of possible scenarios (981 sample sizes x 6 study designs x 4 levels of
 386 endemicity x 3 FEC methods x 3 STH species = 211,896 scenarios), and the different output
 387 parameters (failure rate, $prob_{reduced}$ and $cost_{total}$), we first illustrate the performance of only
 388 KK-based survey designs in areas that are low endemic for hookworm (mean FEC = 3.7 EPG) (**Fig**
 389 **4**). **Fig 4A** shows that the failure rate, i.e., the risk of observing fewer than 50 egg-positive
 390 individuals at baseline, is high (> 25%) when < 250 subjects were enrolled. For sample sizes of
 391 about 250 to 750 subjects, the failure rate was lower for a $SS_{1 \times 2 / 1 \times 2}$ survey design compared to
 392 the other survey designs, as duplicate KK results in higher sensitivity for detecting at least one egg
 393 in the baseline samples. To reduce the failure rate to < 1%, at least 440 individuals needed to be
 394 enrolled for a $SS_{1 \times 2 / 1 \times 2}$, while this was at least 690 for the other survey designs. For the *NS* and
 395 *SSR* survey designs, the probability of correctly detecting reduced drug efficacy ($prob_{reduced}$)
 396 increased with the number of individuals enrolled (**Fig 4B**) but varied between these survey
 397 designs. For example, when 700 individuals were recruited the $prob_{reduced}$ equalled 84% for
 398 $NS_{1 \times 1 / 1 \times 2}$, 79% for $NS_{1 \times 1 / 1 \times 1}$, 71% for $SSR_{1 \times 1 / 1 \times 2}$ and 67% for $SSR_{1 \times 1 / 1 \times 1}$. For *SS* surveys,
 399 $prob_{reduced}$ did not increase beyond 6% and decreased again with increasing sample sizes over

400 400, which is driven by the increasingly precise (due to sample size) but systematically
401 overestimated drug efficacy (due to regression towards the mean).

402

403 Regarding the total survey costs (**Fig 4C**), $NS_{1 \times 1 / 1 \times 2}$, $NS_{1 \times 1 / 1 \times 1}$ and $SS_{1 \times 2 / 1 \times 2}$ were the most

404 expensive survey designs, while the two SSR designs and $SS_{1 \times 1 / 1 \times 1}$ were the cheapest. For

405 example, when enrolling 700 individuals, the mean total survey cost ($cost_{total}$) was 5,853 US\$ for

406 $NS_{1 \times 1 / 1 \times 2}$, 4,861 US\$ for $NS_{1 \times 1 / 1 \times 1}$ and 4,017 US\$ for $SS_{1 \times 2 / 1 \times 2}$, while for the two SSR designs

407 and $SS_{1 \times 1 / 1 \times 1}$ the mean $cost_{total}$ ranged from 2,737 US\$ to 3,157 US\$. To determine the most

408 cost-efficient survey design, we plotted $prob_{reduced}$ as a function of mean $cost_{total}$ (**Fig 4D**) and

409 verified which survey design maximized the $prob_{reduced}$ for a given mean $cost_{total}$. For survey

410 budgets of up to 2,600 US\$, either SSR survey design maximized the $prob_{reduced}$ (lines overlay

411 each other), while for survey budgets between 2,600 US\$ and 4,200 US\$, $NS_{1 \times 1 / 1 \times 1}$ yielded the

412 highest $prob_{reduced}$. When more budget was available, NS surveys were more cost-efficient. For

413 budgets between 4,200 and 5,200 US\$, $NS_{1 \times 1 / 1 \times 1}$ resulted in the highest $prob_{reduced}$, whereas a

414 budget of 5,200 US\$ and above, the NS survey variant deploying a duplicate KK at follow-up

415 ($NS_{1 \times 1 / 1 \times 2}$) was the most cost-efficient. To achieve a <20% risk of falsely concluding adequate

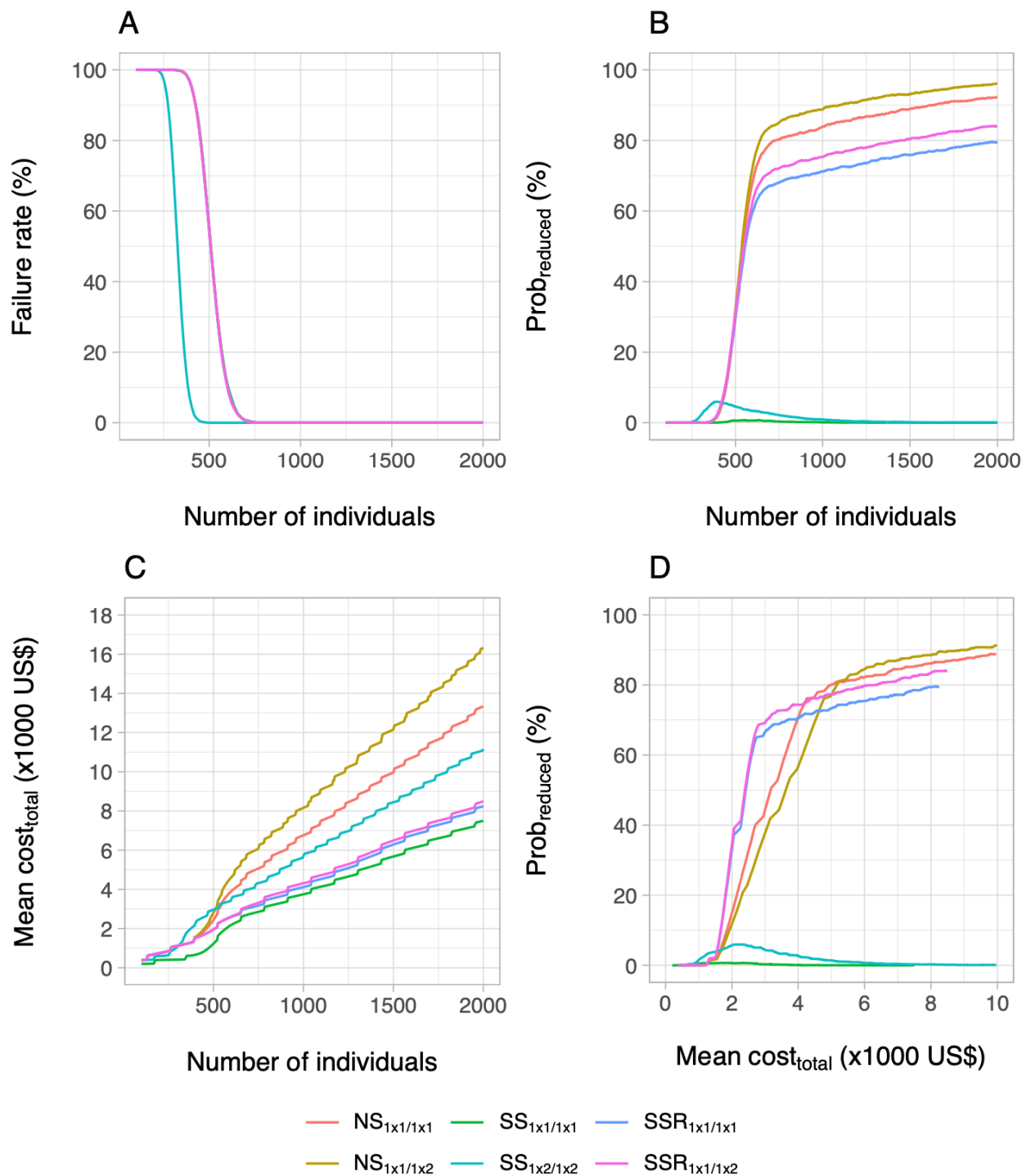
416 drug efficacy ($prob_{reduced} \geq 80\%$), we can see from **Fig 4D** that $NS_{1 \times 1 / 1 \times 1}$ achieves this

417 $prob_{reduced}$ at the lowest cost (red line; mean $cost_{total} = 5,000$ US\$), with the $NS_{1 \times 1 / 1 \times 2}$ as a

418 close runner-up (beige line; mean $cost_{total} = 5,200$ US\$). For a target value of $prob_{reduced} \geq$

419 82%), $NS_{1 \times 1 / 1 \times 2}$ was always the most cost-efficient survey design.

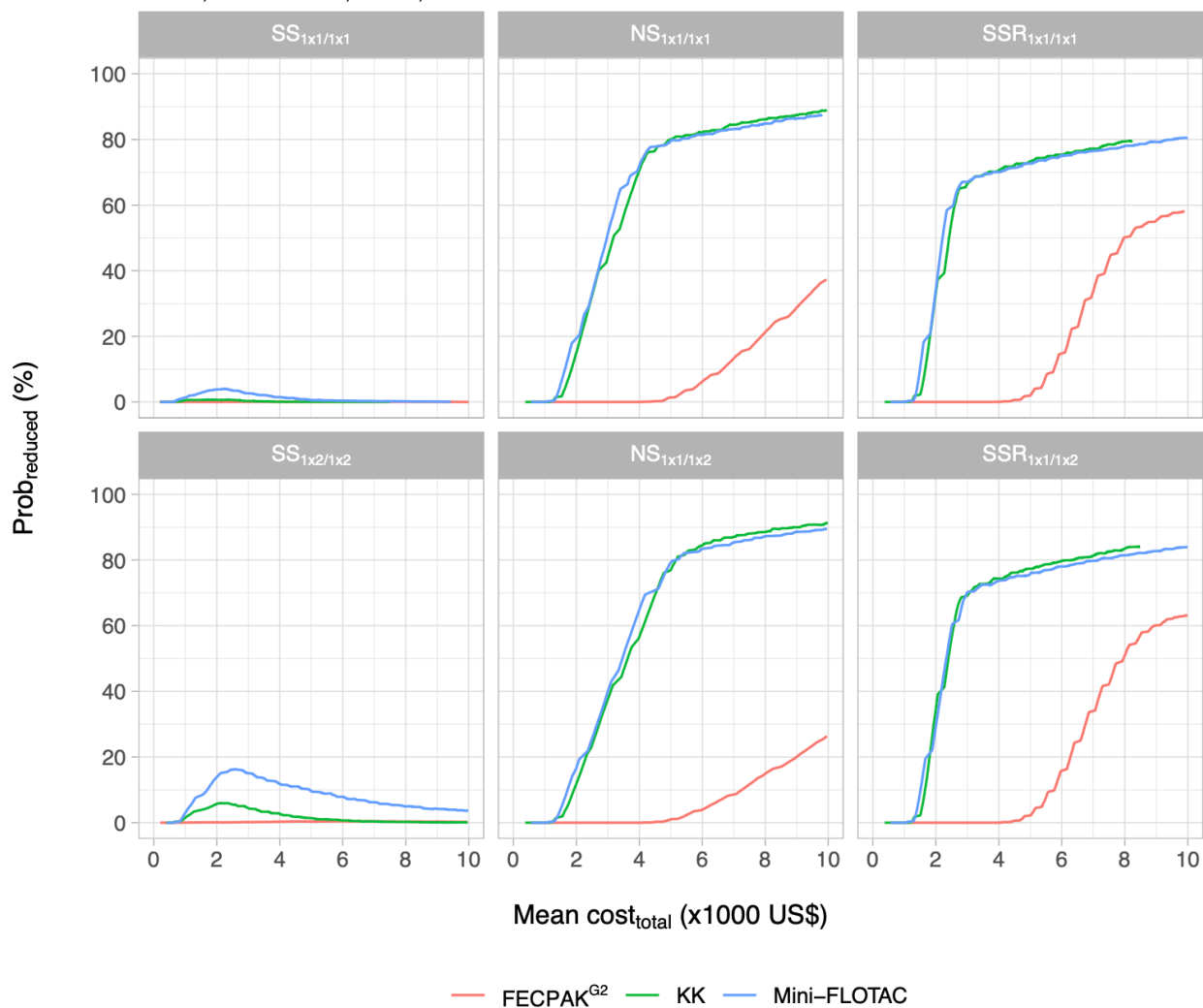
420 **Fig 4: The failure rate, the probability of correctly concluding reduced drug efficacy and the total survey cost across**
 421 **six survey designs.** This figure shows the impact of the survey design and sample size on the failure rate (**Panel A**),
 422 probability of correctly detecting truly reduced efficacy ($prob_{reduced}$; **Panel B**) and the mean total survey cost
 423 ($cost_{total}$; **Panel C**). To gain more insights into the most cost-efficient survey design, the probability of correctly
 424 detecting reduced drug efficacy $prob_{reduced}$ was plotted as a function of the mean $cost_{total}$ (**Panel D**). For each of
 425 the four panels, we only consider the use of Kato-Katz in areas with low levels of hookworm infection (mean FEC = 3.7
 426 EPG). NS = no selection; SS = screen and select; SSR = screen, select, and retest. Note, for panel A, all survey designs
 427 other than SS1x2/1x2 are identical to SSR1x1/1x2.



428

429 Second, we explored the impact of the different FEC methods across the six survey designs for
430 hookworms in the same endemicity level as above (**Fig 5**). Generally, deploying Mini-FLOTAC and
431 FECPAK^{G2} did not greatly improve the *prob_{reduced}* for *SS* survey designs. For the other survey
432 designs, Mini-FLOTAC and KK achieved were equally cost-efficient (lines are close to each other).
433 For Mini-FLOTAC, the cheapest survey design to obtain a *prob_{reduced}* $\geq 80\%$ was an *NS*_{1x1/1x2}
434 survey based on 495 individuals, at a cost of 5,246 US\$ (**S6 Fig**). For KK, this was *NS*_{1x1/1x1} based
435 on 730 individuals at a cost of 4,987 US\$ (**Fig 4**). For FECPAK^{G2}, the *prob_{reduced}* remained below
436 85.2%, even when both sample size (2,000) and available budget (27,140 US\$) were maximized
437 (see **S7 Fig** for details on the impact of sample size).

438 **Fig 5. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for three FEC**
 439 **methods across six survey designs.** This figure plots the probability of correctly identifying reduced therapeutic
 440 efficacy ($prob_{reduced}$) as a function of the mean total survey costs ($cost_{total}$) for the three different FEC methods
 441 (Kato-Katz thick smear (KK), Mini-FLOTAC and FECPAK^{G2}; colored lines) and six survey designs (different panels). For
 442 each panel, we only consider areas that are low endemic for hookworm (mean FEC = 3.7 EPG). NS = no selection; SS
 443 = screen and select; SSR = screen, select, and retest.



444

445 When redrawing **Figs 4 and 5** for the other two STH species at the corresponding low endemicity

446 level (*A. lumbricoides*: mean FEC = 9.6 EPG, **Figs S2 – 3**; *T. trichiura*: mean FEC = 2.8 EPG, **Figs S4 –**

447 **5**), we noted three important differences compared to hookworm. First, the risk for a failed survey

448 was remarkably lower for *A. lumbricoides*. While the failure rate is 8.5% when 250 subjects are

449 enrolled for a survey (SS_{1x2/1x2}) targeting *A. lumbricoides* (**Fig S2A**), this was 98.7% and 97.4% for

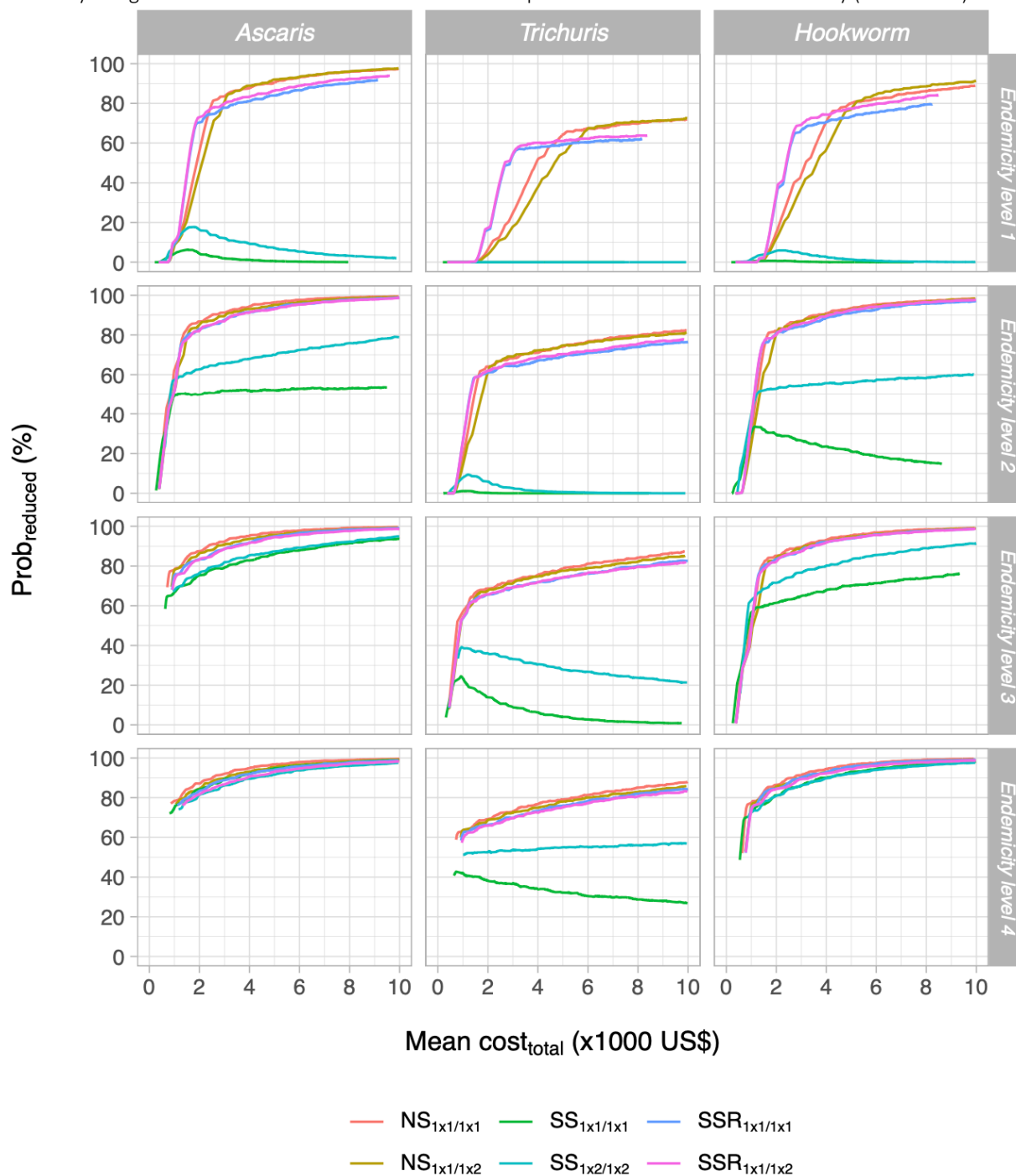
450 *T. trichiura* (**Fig S4A**) and hookworms (**Fig 4A**), respectively. As a consequence of this, the mean

451 $cost_{total}$ and the sample size at which $prob_{reduced} \geq 80\%$ was lower compared to the other
452 STHs (*A. lumbricoides*: **Fig S2D**: $NS_{1 \times 1 / 1 \times 1}$ at mean $cost_{total} = 2,522$ US\$; *T. trichiura*: **Fig S4D**:
453 $NS_{1 \times 1 / 1 \times 2}$ at mean $cost_{total} = 19,628$ US\$; Hookworm: **Fig 4D**: $NS_{1 \times 1 / 1 \times 1}$ at mean $cost_{total} =$
454 4,987 US\$). Second, in contrast to hookworms, where $prob_{reduced}$ for a given budget differed
455 only marginally between Mini-FLOTAC and Kato-Katz thick smear (**Fig 5**), the differences FEC
456 methods were more substantial for *A. lumbricoides* and *T. trichiura*. For *A. lumbricoides* (**Fig S3**),
457 KK provided the highest $prob_{reduced}$ for the same budget, while this was Mini-FLOTAC for *T.*
458 *trichiura* (**Fig S5**). Third, none of the survey designs achieved a $prob_{reduced} \geq 80\%$ for *T. trichiura*,
459 given the maximum simulated sample size of 2,000 individuals (**Fig S4B**).

460
461 In **Fig 6**, we show the impact of pre-treatment endemicity on the probability of correctly identifying
462 reduced drug efficacy based on KK. When surveys were conducted in higher levels of endemicity,
463 a higher $prob_{reduced}$ was obtained for the same budget. Although this was observed for all three
464 STH species and all six survey designs, this increase was most distinct for *SS* survey designs. This
465 was to be expected as the bias due to regression towards the mean in *SS* survey designs is known
466 to decrease with higher infection levels. Although *SS* survey designs rarely resulted in a correct
467 detection of a truly reduced therapeutic drug efficacy when endemicity levels were low (top row
468 panels **Fig 6**), they almost reach the highest $prob_{reduced}$ at a cost of the cheapest survey design
469 at the highest levels of endemicity for both *A. lumbricoides* and hookworms (bottom row panels
470 **Fig 6**). For *T. trichiura*, the performance of *SS* survey designs remained relatively poor, which is
471 logical as the regression towards the mean is expected to be higher when true drug efficacy is
472 lower (45% for *T. trichiura* in the simulations). This figure also highlights a shift in the most cost-

473 efficient survey design: while at low level of endemicity, the most cost-efficient survey design
474 depends on the available funds, for higher endemicities, only the $NS_{1 \times 1 / 1 \times 1}$ survey design
475 maximizes the $prob_{reduced}$ for any available budget.

476 Fig 6. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for six survey
 477 designs across four levels of endemicity when deploying Kato-Katz. This figure plots the probability of correctly
 478 identifying reduced therapeutic efficacy ($prob_{reduced}$) as a function of the mean total survey costs ($cost_{total}$) across
 479 six survey designs for the three soil-transmitted helminth species and four levels of endemicity (see Table 1).



480

481

482 In **Table 4**, we provide the sample size and mean $cost_{total}$ for those survey designs that detected
 483 reduced efficacy with $prob_{reduced} \sim 80\%$ at the lowest cost for each of the different STH species.
 484 Generally, this table confirms that the $NS_{1 \times 1 / 1 \times 1}$ survey design in combination with KK was the
 485 most cost-efficient choice to assess therapeutic drug efficacy in all scenarios of STH species and
 486 endemicity. Only when surveys were conducted in areas where endemicity of *T. trichiura*
 487 infections were low, the $NS_{1 \times 1 / 1 \times 2}$ survey combined with Mini-FLOTAC was more cost-efficient.
 488

489 **Table 4. The most cost-efficient survey design and FEC method to monitor the therapeutic drug efficacy against STHs.**

STH species	Mean FEC (EPG)	Survey design	FEC method	Sample size	Mean $cost_{total}$ (US\$)
<i>A. lumbricoides</i>					
	9.6	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	380	2,522
	85.2	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	165	1,358
	360.0	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	145	1,282
	2195.5	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	135	1,243
<i>T. trichiura</i>					
	2.8	$NS_{1 \times 1 / 1 \times 2}$	Mini-FLOTAC	1,680	17,313
	12.9	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	1,215	8,134
	49.7	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	830	5,740
	124.7	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	790	5,523
Hookworms					
	3.7	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	730	4,987
	23.7	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	245	1,666
	61.7	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	185	1,437
	210.3	$NS_{1 \times 1 / 1 \times 1}$	Kato-Katz	160	1,340

490

491 Discussion

492 In this paper, we present new evidence-based recommendations for cost-efficient monitoring of
 493 therapeutic drug efficacy against STH, using a simulation framework that captures important
 494 interactions between STH epidemiology (variability in egg counts due to various sources),
 495 diagnostic test performance (species- and method-dependent egg recovery and count variability),

496 survey design (bias and accuracy that change with endemicity) and operational costs (which
497 change with endemicity, diagnostic method and survey design). With this framework, we address
498 the challenge of minimizing operational costs of STH monitoring in resource-limited settings
499 without jeopardizing the quality of the decision-making. We performed an in-depth analysis of the
500 operational costs to process one sample for three FEC methods (KK, Mini-FLOTAC and FECPAK^{G2})
501 based on the time-to-result and an itemized cost assessment. Next, we simulated how the
502 probability of correctly detecting reduced therapeutic drug efficacy depends on different FEC
503 methods and survey designs, accounting for sources of variation in egg counts as quantified based
504 on several STH datasets. Finally, we integrated the outcome of the in-depth cost-assessment into
505 the simulation study and determined the most cost-efficient survey design to detect presence of
506 reduced drug efficacy across different scenarios of STH endemicity. Overall, we confirm that KK is
507 the best FEC method to monitor therapeutic drug efficacy, but that the survey design currently
508 recommended by WHO should be updated.

509

510 Single KK is the cheapest and least time-consuming method

511 The mean time and cost for material to process one sample varied from ~8.5 min (single KK) to
512 ~16.5 min (duplicate KK), and from US\$ 1.37 (single KK) to US\$ 1.69 (FECPAK^{G2}). Our study found
513 that a single KK is both the cheapest and least-time consuming of the three FEC methods
514 evaluated. Although a comparison across studies is not straightforward, as differences in
515 laboratory time can be explained by differences in endemicity (laboratory time varies significantly
516 with the number of eggs counted) and level of expertise of laboratory technicians (laboratory time

517 was lower when technicians were more acquainted with the FEC method; data not shown), other
518 researchers generally reported a similar laboratory time for both single ([35: ~9.5 min; [36]; ~11.0
519 min; [37]: ~5.0 min) and duplicate KK ([35]: ~16.6 min). The cost for material estimated in the
520 current study (single KK: US\$ 1.95; duplicate KK: US\$ 2.17), were higher than those reported by
521 Speich *et al.* [35] (single KK: US\$ 0.03; duplicate KK: US\$ 0.04). These differences can be explained
522 by the fact that we included fixed survey costs (0.60 US\$, e.g., gloves and permanent markers). In
523 addition, while Speich *et al* [35] re-used the templates for 50 samples, we opted for single use of
524 materials as there was only limited mesh and cellophane in one kit. For the other two FEC
525 methods, the laboratory time and cost for material were ~13.1 min and US\$ 1.51 for Mini-FLOTAC,
526 and ~13.5 min and US\$ 1.69 for FECPAK^{G2}. However, data on laboratory time and cost for material
527 to compare our results are either scarce (Mini-FLOTAC: 8 – 12 min [38]) or absent (FECPAK^{G2}). It is
528 also important to note that our estimates of laboratory time and material costs did not include the
529 washing of devices, and that we based our costs based on an Ethiopian market in 2020.

530

531 Revision of the WHO guidelines to monitor drug efficacy is warranted

532 WHO currently recommends a selection and screen approach during which a single stool sample
533 is processed by a single KK both at baseline and follow-up [5]. Although our study confirms that KK
534 is the FEC method of choice, it indicates that the recommended survey design will often result in
535 poor decision-making due to overestimation of drug efficacy (because of regression towards the
536 mean) at a relatively high cost. Instead, a KK-based survey design where all children are followed
537 up regardless of their baseline infection status (the “no selection” or *NS* design) should be

538 preferred as it yields unbiased results at the lowest operational cost. The “screen, select, and
539 retest” strategy, where individuals who are egg-positive at baseline are retested based on a new
540 pre-treatment stool sample (the *SSR* design) [21], was found to be somewhat less cost-efficient
541 than the *NS* design. However, as previously discussed [21], because the *SSR* design will yield more
542 egg-positive individuals than a *NS* design based on the same budget, *SSR* could still be considered
543 for study objectives that require a minimal number of eggs or egg-positive individuals, such as
544 genotyping to identify resistance-conferring polymorphisms.

545

546 Identifying the most cost-efficient study design for any programmatic survey

547 In the present study, the laboratory time and cost analysis were used to identify the most cost-
548 efficient design for monitoring drug efficacy, but these analyses can also be used to identify the
549 most cost-effective study design for any other type of survey. Although this concept is not new
550 and has been applied in the past [39,40], the level of detail that we present for each of the different
551 FEC allows for fine-tuned cost-efficiency analysis of any programmatic survey. For instance, this
552 framework would also lend itself well to assess the cost and performance of surveys for decisions
553 about stopping or scaling down preventive chemotherapy against STH and/or assess the potential
554 value and cost-efficiency of (new) diagnostic techniques with different (hopefully better)
555 performance and throughput than FEC methods, but potentially higher reagent costs [41].

556

557 Automated egg counting would further reduce operational costs

558 As highlighted by the present study, egg counting is the most time-consuming step for KK and
559 Mini-FLOTAC (80%). An obvious cost-saving strategy that would further reduce the operational
560 costs is automated egg counting using a scanning/imaging device and artificial intelligence-based
561 egg-recognition software to identify and report egg counts.-A variety of artificial intelligence based
562 digital pathology (AI-DP) devices are currently being studied [42-46]. However, a complete AI-DP
563 device is currently not commercially available, despite the successful examples for other parasitic
564 infections (malaria: CellsCheck, <http://www.biosynex.com>; Loa Loa: [47]). At the time of writing,
565 FECPAK^{G2} was probably the most advanced, but automated egg-recognition on the created images
566 by existing STH egg-recognition software has proven difficult or impossible (S5 Info of Cools et al.
567 [30]). In addition, our study highlighted that due to its poor diagnostic performance [30], FECPAK^{G2}
568 is not recommended to monitor therapeutic drug efficacy in STH control programs. Despite these
569 challenges, there are ongoing investments around each of the FEC methods to progress towards
570 a complete point-of-care platform with automated egg counting and built-in data analysis [43,46].
571

572 Strengths and limitations

573 This is the first comprehensive study that compares the operational costs between the most-used
574 FEC methods in STH surveys. It is important to note that the estimated costs are institute- and
575 context-specific, and hence the reported values should not be interpreted as absolute. However,
576 a major strength of our framework is that assumptions about costs can be easily adapted to
577 represent particular settings. Because our framework aims to compare different survey designs

578 and diagnostic methods, we do not consider costs that can be reasonably assumed to be the
579 similar across different survey designs and different FEC methods: salary for senior staff to
580 supervise the field activities, report, and analyze the data; power supplies; laboratory rent; per
581 diems for days when work is not possible (e.g., weekends); and time required to set up and clean
582 laboratories and inform schools and local health authorities prior to surveys. A possibly relevant
583 simplifying assumption that we made is that survey teams work constantly without any break. This
584 may have led to a slight overestimation of the performance per cost (**Fig 5**) for each of the different
585 FEC methods (more breaks because of manual egg counting) compared to FECKPAK^{G2}. Finally, it is
586 important to highlight that each of the trials were conducted by well-trained teams (and hence
587 the laboratory time for a less experienced team might be underestimated), and that we assumed
588 that no individuals would be lost to follow-up. Theoretically, each of the aforementioned factors
589 could be included in our simulation framework, although we do not expect that the presented
590 relative rankings of FEC methods and survey designs would be affected by the inclusion of this
591 additional real-life complexity. However, some of these aspects, like laboratory infrastructure, will
592 have to be considered when comparing FEC methods with other diagnostic techniques such as
593 quantitative polymerase chain reaction.

594

595 Conclusion

596 We confirm that Kato-Katz is the FEC method of choice for assessing drug efficacy, but that the
597 current WHO-recommended screen and select survey design should be replaced by a no-screen
598 survey design. Our detailed analysis of laboratory time and material costs allows for a cost-

599 efficiency assessment of other FEC-based programmatic decisions as well, like decisions regarding
600 stopping and scaling down of preventive chemotherapy. In addition, the flexible and holistic design
601 of our simulation framework allows further study of alternative diagnostic techniques that aim to
602 further decrease operational costs while maintaining or improving the quality of decision-making,
603 like automated egg counting.

604

605 **References**

- 606 1. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted
607 life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries
608 and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.
609 Lancet. 2018 Nov 10;392(10159):1859-1922. doi: 10.1016/S0140-6736(18)32335-3. Erratum in:
610 Lancet. 2019 Jun 22;393(10190):e44.
- 611 2. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted
612 life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries
613 and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016.
614 Lancet. 2017 Sep 16;390(10100):1260-1344. doi: 10.1016/S0140-6736(17)32130-X. Erratum in:
615 Lancet. 2017 Oct 28;390(10106):e38.
- 616 3. World Health Organization. Preventive chemotherapy to control soil-transmitted helminth
617 infections in at-risk population groups. 2017, World Health Organization; Geneva, Switzerland.
- 618 4. World Health Organization. Soil-transmitted helminthiases: eliminating soil-transmitted
619 helminthiases as a public health problem in children: progress report 2001-2010 and strategic plan
620 2011-2020. 2012: Geneva, Switzerland.
- 621 5. World Health Organization. Assessing the efficacy of anthelmintic drugs against
622 schistosomiasis and soil-transmitted helminthiases. 2013, World Health Organization Geneva.
- 623 6. World Health Organization. 2030 targets for soil-transmitted helminthiases control
624 programmes. 2020. World Health Organization; Geneva, Switzerland.

- 625 7. Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, Rollinson D, Utzinger J.
626 Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted
627 helminth diagnosis in epidemiological surveys. *Parasit Vectors*. 2010 Aug 14;3:71. doi:
628 10.1186/1756-3305-3-71.
- 629 8. Leta GT, French M, Dorny P, Vercruysse J, Levecke B. Comparison of individual and pooled
630 diagnostic examination strategies during the national mapping of soil-transmitted helminths and
631 *Schistosoma mansoni* in Ethiopia. *PLoS Negl Trop Dis*. 2018 Sep 10;12(9):e0006723. doi:
632 10.1371/journal.pntd.0006723.
- 633 9. Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruysse J, Utzinger J, Rinaldi L. The Mini-
634 FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals.
635 *Nat Protoc*. 2017 Sep;12(9):1723-1732. doi: 10.1038/nprot.2017.067.
- 636 10. Bekana T, Mekonnen Z, Zeynudin A, Ayana M, Getachew M, Vercruysse J, Levecke B.
637 Comparison of Kato-Katz thick-smear and McMaster egg counting method for the assessment of
638 drug efficacy against soil-transmitted helminthiasis in school children in Jimma Town, Ethiopia.
639 *Trans R Soc Trop Med Hyg*. 2015 Oct;109(10):669-71. doi: 10.1093/trstmh/trv073.
- 640 11. Ayana M, Vlamincck J, Cools P, Ame S, Albonico M, Dana D, et al. Modification and
641 optimization of the FECPAKG2 protocol for the detection and quantification of soil-transmitted
642 helminth eggs in human stool. *PLoS Negl Trop Dis*. 2018 Oct 15;12(10):e0006655. doi:
643 10.1371/journal.pntd.0006655.
- 644 12. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique
645 in *Schistosomiasis mansoni*. *Rev Inst Med Trop Sao Paulo*. 1972 Nov-Dec;14(6):397-400.
- 646 13. TechionGroup.com. 2017; Available from: <https://www.techion.com/FECPAKG2>.

- 647 14. Rashid MH, Stevenson MA, Waenga S, Mirams G, Campbell AJD, Vaughan JL, Jabbar A.
648 Comparison of McMaster and FECPAKG2 methods for counting nematode eggs in the faeces of
649 alpacas. *Parasit Vectors*. 2018 May 2;11(1):278. doi: 10.1186/s13071-018-2861-1.
- 650 15. Cools P, Vlaminck J, Albonico M, Ame S, Ayana M, José Antonio BP, et al. Diagnostic
651 performance of a single and duplicate Kato-Katz, Mini-FLOTAC, FECPAKG2 and qPCR for the
652 detection and quantification of soil-transmitted helminths in three endemic countries. *PLoS Negl*
653 *Trop Dis*. 2019 Aug 1;13(8):e0007446. doi: 10.1371/journal.pntd.0007446.
- 654 16. Levecke B, Cools P, Albonico M, Ame S, Angebault C, Ayana M, et al. Identifying thresholds
655 for classifying moderate-to-heavy soil-transmitted helminth intensity infections for FECPAKG2,
656 McMaster, Mini-FLOTAC and qPCR. *PLoS Negl Trop Dis*. 2020 Jul 2;14(7):e0008296. doi:
657 10.1371/journal.pntd.0008296.
- 658 17. Levecke B, Anderson RM, Berkvens D, Charlier J, Devleeschauwer B, Speybroeck N,
659 Vercruyse J, Van Aelst S. Mathematical inference on helminth egg counts in stool and its
660 applications in mass drug administration programmes to control soil-transmitted helminthiasis in
661 public health. *Adv Parasitol*. 2015 Mar;87:193-247. doi: 10.1016/bs.apar.2015.01.001.
- 662 18. Gass KM. Rethinking the serological threshold for onchocerciasis elimination. *PLoS Negl*
663 *Trop Dis*. 2018 Mar 15;12(3):e0006249. doi: 10.1371/journal.pntd.0006249.
- 664 19. Coffeng LE, Stolk WA, Golden A, de Los Santos T, Domingo GJ, de Vlas SJ. Predictive Value
665 of Ov16 Antibody Prevalence in Different Subpopulations for Elimination of African
666 Onchocerciasis. *Am J Epidemiol*. 2019 Sep 1;188(9):1723-1732. doi: 10.1093/aje/kwz109.

- 667 20. Coffeng LE, Malizia V, Vegvari C, Cools P, Halliday KE, Levecke B, Mekonnen Z, Gichuki PM,
668 Sayasone S, Sarkar R, Shaali A, Vlaminck J, Anderson RM, de Vlas SJ. Impact of Different Sampling
669 Schemes for Decision Making in Soil-Transmitted Helminthiasis Control Programs. *J Infect Dis.*
670 2020 Jun 11;221(Suppl 5):S531-S538. doi: 10.1093/infdis/jiz535.
- 671 21. Coffeng LE, Levecke B, Hattendorf J, Walker M, Denwood MJ. Survey Design to Monitor
672 Drug Efficacy for the Control of Soil-Transmitted Helminthiasis and Schistosomiasis. *Clin Infect Dis.*
673 2021 Jun 14;72(Suppl 3):S195-S202. doi: 10.1093/cid/ciab196.
- 674 22. Coffeng LE, Le Rutte EA, Munoz J, Adams E, de Vlas SJ. Antibody and Antigen Prevalence as
675 Indicators of Ongoing Transmission or Elimination of Visceral Leishmaniasis: A Modeling Study.
676 *Clin Infect Dis.* 2021 Jun 14;72(Suppl 3):S180-S187. doi: 10.1093/cid/ciab210.
- 677 23. Levecke B, Coffeng LE, Hanna C, Pullan RL, Gass KM. Assessment of the required
678 performance and the development of corresponding program decision rules for neglected tropical
679 diseases diagnostic tests: Monitoring and evaluation of soil-transmitted helminthiasis control
680 programs as a case study. *PLoS Negl Trop Dis.* 2021 Sep 14;15(9):e0009740. doi:
681 10.1371/journal.pntd.0009740.
- 682 24. Vlaminck J, Cools P, Albonico M, Ame S, Ayana M, et al. Comprehensive evaluation of stool-
683 based diagnostic methods and benzimidazole resistance markers to assess drug efficacy and
684 detect the emergence of anthelmintic resistance: A Starworms study protocol. *PLoS Negl Trop Dis.*
685 2018 Nov 2;12(11):e0006912. doi: 10.1371/journal.pntd.0006912.
- 686 25. Vlaminck J, Cools P, Albonico M, Ame S, Ayana M, Cringoli G, et al. Therapeutic efficacy of
687 albendazole against soil-transmitted helminthiasis in children measured by five diagnostic
688 methods. *PLoS Negl Trop Dis.* 2019 Aug 1;13(8):e0007471. doi: 10.1371/journal.pntd.0007471.

- 689 26. R Core Team (2022). R: A language and environment for statistical computing. R
690 Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 691 27. Vlaminc J, Cools P, Albonico M, Ame S, Chanthapaseuth T, Viengxay V, Do Trung D, et al.
692 Piloting a surveillance system to monitor the global patterns of drug efficacy and the emergence
693 of anthelmintic resistance in soil-transmitted helminth control programs: a Starworms study
694 protocol. *Gates Open Res.* 2020 Mar 10;4:28. doi: 10.12688/gatesopenres.13115.1.
- 695 28. Denwood MJ, Love S, Innocent GT, Matthews L, McKendrick IJ, Hillary N, et al. Quantifying
696 the sources of variability in equine faecal egg counts: implications for improving the utility of the
697 method. *Vet Parasitol.* 2012 Aug 13;188(1-2):120-6. doi: 10.1016/j.vetpar.2012.03.005. Epub
698 2012 Mar 13.
- 699 29. Knopp S, Mohammed KA, Speich B, Hattendorf J, Khamis IS, Khamis AN, et al. Albendazole
700 and mebendazole administered alone or in combination with ivermectin against *Trichuris*
701 *trichiura*: a randomized controlled trial. *Clin Infect Dis.* 2010 Dec 15;51(12):1420-8. doi:
702 10.1086/657310.
- 703 30. Cools P, Vlaminc J, Albonico M, Ame S, Ayana M, José Antonio BP, et al. Diagnostic
704 performance of a single and duplicate Kato-Katz, Mini-FLOTAC, FECPAKG2 and qPCR for the
705 detection and quantification of soil-transmitted helminths in three endemic countries. *PLoS Negl*
706 *Trop Dis.* 2019 Aug 1;13(8):e0007446. doi: 10.1371/journal.pntd.0007446.
- 707 31. Barda B, Cajal P, Villagran E, Cimino R, Juarez M, Krolewiecki A, et al. Mini-FLOTAC, Kato-
708 Katz and McMaster: three methods, one goal; highlights from north Argentina. *Parasit Vectors.*
709 2014 Jun 14;7:271. doi: 10.1186/1756-3305-7-271.

- 710 32. Nikolay B, Mwandawiro CS, Kihara JH, Okoyo C, Cano J, Mwanje MT, et al. Understanding
711 Heterogeneity in the Impact of National Neglected Tropical Disease Control Programmes:
712 Evidence from School-Based Deworming in Kenya. *PLoS Negl Trop Dis*. 2015 Sep
713 30;9(9):e0004108. doi: 10.1371/journal.pntd.0004108.
- 714 33. Levecke B, Speybroeck N, Dobson RJ, Vercruysse J, Charlier J. Novel insights in the fecal egg
715 count reduction test for monitoring drug efficacy against soil-transmitted helminths in large-scale
716 treatment programs. *PLoS Negl Trop Dis*. 2011 Dec;5(12):e1427. doi:
717 10.1371/journal.pntd.0001427.
- 718 34. R-package; Available from: <http://ku-awdc.github.io/eggSim>.
- 719 35. Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, Rollinson D, Utzinger J.
720 Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted
721 helminth diagnosis in epidemiological surveys. *Parasit Vectors*. 2010 Aug 14;3:71. doi:
722 10.1186/1756-3305-3-71.
- 723 36. Kure A, Mekonnen Z, Dana D, Bajiro M, Ayana M, Vercruysse J, Levecke B. Comparison of
724 individual and pooled stool samples for the assessment of intensity of *Schistosoma mansoni* and
725 soil-transmitted helminth infections using the Kato-Katz technique. *Parasit Vectors*. 2015 Sep
726 24;8:489. doi: 10.1186/s13071-015-1101-1.
- 727 37. Leta GT, French M, Dorny P, Vercruysse J, Levecke B. Comparison of individual and pooled
728 diagnostic examination strategies during the national mapping of soil-transmitted helminths and
729 *Schistosoma mansoni* in Ethiopia. *PLoS Negl Trop Dis*. 2018 Sep 10;12(9):e0006723. doi:
730 10.1371/journal.pntd.0006723.

- 731 38. Cringoli G, Maurelli MP, Levecke B, Bosco A, Vercruyse J, Utzinger J, Rinaldi L. The Mini-
732 FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals.
733 Nat Protoc. 2017 Sep;12(9):1723-1732. doi: 10.1038/nprot.2017.067. Epub 2017 Aug 3.
- 734 39. Assefa LM, Crellen T, Kepha S, Kihara JH, Njenga SM, Pullan RL, Brooker SJ. Diagnostic
735 accuracy and cost-effectiveness of alternative methods for detection of soil-transmitted helminths
736 in a post-treatment setting in western Kenya. PLoS Negl Trop Dis. 2014 May 8;8(5):e2843. doi:
737 10.1371/journal.pntd.0002843.
- 738 40. Sturrock HJ, Gething PW, Clements AC, Brooker S. Optimal survey designs for targeting
739 chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-
740 efficiency of sampling. Am J Trop Med Hyg. 2010 Jun;82(6):1079-87. doi: 10.4269/ajtmh.2010.09-
741 0702.
- 742 41. Kazienga A, Coffeng LE, de Vlas SJ, Levecke B. Two-stage lot quality assurance sampling
743 framework for monitoring and evaluation of neglected tropical diseases, allowing for imperfect
744 diagnostics and spatial heterogeneity. PLoS Negl Trop Dis. 2022 Apr 8;16(4):e0010353. doi:
745 10.1371/journal.pntd.0010353.
- 746 42. Larsson, J. and R. Hedberg, Development of machine learning models for object
747 identification of parasite eggs using microscopy. 2020.
- 748 43. Dacal E, Bermejo-Peláez D, Lin L, Álamo E, Cuadrado D, Martínez Á, Mousa A, et al. Mobile
749 microscopy and telemedicine platform assisted by deep learning for the quantification of *Trichuris*
750 *trichiura* infection. PLoS Negl Trop Dis. 2021 Sep 7;15(9):e0009677. doi:
751 10.1371/journal.pntd.0009677.

- 752 44. Ephraim RK, Duah E, Cybulski JS, Prakash M, D'Ambrosio MV, Fletcher DA, et al. Diagnosis
753 of *Schistosoma haematobium* infection with a mobile phone-mounted Foldscope and a reversed-
754 lens CellScope in Ghana. *Am J Trop Med Hyg.* 2015 Jun;92(6):1253-6. doi: 10.4269/ajtmh.14-0741.
- 755 45. Cringoli G, Amadesi A, Maurelli MP, Celano B, Piantadosi G, Bosco A, et al. The Kubic
756 FLOTAC microscope (KFM): a new compact digital microscope for helminth egg counts.
757 *Parasitology.* 2021 Apr;148(4):427-434. doi: 10.1017/S003118202000219X.
- 758 46. Ward P, Dahlberg P, Lagatie O, Larsson J, Tynong A, Vlaminck J, Zumpe M, Ame S, Ayana
759 M, Khieu V, Mekonnen Z, Odiere M, Yohannes T, Van Hoecke S, Levecke B, Stuyver LJ. Affordable
760 artificial intelligence-based digital pathology for neglected tropical diseases: A proof-of-concept
761 for the detection of soil-transmitted helminths and *Schistosoma mansoni* eggs in Kato-Katz stool
762 thick smears. *PLoS Negl Trop Dis.* 2022 Jun 17;16(6):e0010500. doi:
763 10.1371/journal.pntd.0010500.
- 764 47. D'Ambrosio MV, Bakalar M, Bennuru S, Reber C, Skandarajah A, Nilsson L, et al. Point-of-
765 care quantification of blood-borne filarial parasites with a mobile phone microscope. *Sci Transl*
766 *Med.* 2015 May 6;7(286):286re4. doi: 10.1126/scitranslmed.aaa3480.
- 767

768 Acknowledgements

769 First and foremost, the authors would like to express their gratitude towards all children, their
770 parents, the school teachers and principals that participated in this study. Second, we wish to
771 specifically thank all the people that provided the necessary laboratory and logistic support in each
772 of the four different sampling sites. This work would not have been possible without their willful
773 participation and assistance.

774 First and foremost, the authors would like to express their gratitude towards all children, their
775 parents, the school teachers and principals that participated in this study. Second, we wish to
776 specifically thank all the people that provided the necessary laboratory and logistic support in each
777 of the four different sampling sites. This work would not have been possible without their willful
778 participation and assistance.

779

Supporting Information Legends

Info S1. Summary of the standard operating procedures to time the preparatory steps and the egg counting for three FEC methods.

Info S2. Summary of the standard operating procedures to time the data entry.

Info S3. A detailed itemized cost assessment both to collect and process stool samples and for each FEC method.

Info S4. Methods to estimate the FECs of Mini-FLOTAC and FECPAK^{G2} based on duplicate Kato-Katz.

Info S5. Detailed description and parameterization of the simulation model.

Info S6. The calculation of the total operational costs to monitor drug efficacy

Info S7. Detail information on the calculation of the time-to-result

Fig S1. The range of school mean fecal egg counts for four levels of school prevalence for three soil-transmitted helminths

Fig S2. The failure rate, the probability of correctly concluding reduced drug efficacy and the total survey cost across six survey designs for *Ascaris*. This figure shows the impact of the survey design and sample size on the failure rate (**Panel A**), probability of correctly detecting truly reduced efficacy ($prob_{reduced}$; **Panel B**) and the mean total survey cost ($cost_{total}$; **Panel C**). To gain more insights into the most cost-efficient survey design, the probability of correctly detecting reduced drug efficacy $prob_{reduced}$ was plotted as a function of the mean $cost_{total}$ (**Panel D**). For each of the four panels, we only consider the use of Kato-Katz in areas with low levels of *Ascaris* infection (mean FEC = 9.6 EPG). NS = no selection; SS = screen and select; SSR = screen, select, and retest. Note, for panel A, all survey designs other than SS1x2/1x2 are identical to SSR1x1/1x2.

Fig S3. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for three FEC methods across six survey designs for *Ascaris*. This figure plots the probability of correctly identifying reduced therapeutic efficacy ($prob_{reduced}$) against *Ascaris* infections as a function of the mean total survey costs ($cost_{total}$) for the three different FEC methods (Kato-Katz thick smear (KK), Mini-FLOTAC and FECPAK^{G2}; colored lines) and six survey designs (different panels). For each panel, we only consider areas that are low endemic for *Ascaris* (mean FEC = 9.6 EPG). NS = no selection; SS = screen and select; SSR = screen, select, and retest.

Fig S4. The failure rate, the probability of correctly concluding reduced drug efficacy and the total survey cost across six survey designs for *Trichuris*. This figure shows the impact of the survey design and sample size on the failure rate (**Panel A**), probability of correctly detecting truly reduced efficacy ($prob_{reduced}$; **Panel B**) and the mean total survey cost ($cost_{total}$; **Panel C**). To gain more insights into the most cost-efficient survey design, the probability of correctly detecting reduced drug efficacy $prob_{reduced}$ was plotted as a function of the mean $cost_{total}$ (**Panel D**). For each of the four panels, we only consider the use of Kato-Katz in areas with low levels of *Trichuris* infection (mean FEC = 2.8 EPG). NS = no selection; SS = screen and select; SSR = screen, select, and retest. Note, for panel A, all survey designs other than SS1x2/1x2 are identical to SSR1x1/1x2.

Fig S5. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for three FEC methods across six survey designs for *Trichuris*. This figure plots the probability of correctly identifying reduced therapeutic efficacy ($prob_{reduced}$) against *Trichuris* infections as a function of the mean total survey costs ($cost_{total}$) for the three different FEC methods (Kato-Katz thick smear (KK), Mini-FLOTAC and FECPAK^{G2}; colored lines) and six survey designs (different panels). For each panel, we only consider areas that are low endemic for *Trichuris* (mean FEC = 2.8 EPG). NS = no selection; SS = screen and select; SSR = screen, select, and retest.

Fig S6. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for six survey designs across four levels of endemicity when deploying Mini-FLOTAC. This figure plots the probability of correctly identifying reduced therapeutic efficacy ($prob_{reduced}$) as a function of the mean total survey costs ($cost_{total}$) across six survey designs for the three soil-transmitted helminth species and four levels of endemicity (see **Table 1**).

Fig S7. The probability of correctly detecting presence of reduced drug efficacy and the total survey cost for six survey designs across four levels of endemicity when deploying FECPAK^{G2}. This figure plots the probability of correctly identifying reduced therapeutic efficacy ($prob_{reduced}$) as a function of the mean total survey costs ($cost_{total}$) across six survey designs for the three soil-transmitted helminth species and four levels of endemicity (see **Table 1**).