1 2 3 4 5	Overestimation of School-Based Deworming Coverage Resulting from School-Based Reporting
6	William Sheahan ^{1¶*#a} , Roy Anderson ² , Kumudha Aruldas ³ , Euripide Avokpaho ⁴ , Sean Galagan5, Jeanne
7	Goodman ⁵ , Parfait Houngbegnon ⁴ , Gideon John Israel ³ , Venkateshprabhu Janagaraj ³ , Saravanakumar
8	Puthupalayam Kaliappan³, Arianna Rubin Means⁵, Chloe Morozoff⁵, Emily Pearman⁵, Rohan Michael
9	Ramesh ³ , Amy Roll ⁵ , Alex Schaefer ⁵ , Sitara S. R. Ajjampur ^{3&} , Robin Bailey ^{6&} , Moudachirou Ibikounlé ^{4&} ,
10	Khumbo Kalua ^{6&} , Adrian J.F. Luty ^{7&} , Rachel Pullan ^{8&} , Judd L. Walson ^{5,9,10} , Kristjana Hrönn
11	Ásbjörnsdóttir ^{5,11}
12 13	
13 14 15	1 Malaria and Neglected Tropical Diseases, PATH, Seattle, Washington, United States of America
16 17	2 School of Public Health, Faculty of Medicine, Imperial College, London, United Kingdom
18 19	3 The Wellcome Trust Research Laboratory, Division of Gastrointestinal Sciences, Christian Medical College, Vellore, India
20 21 22	4 Institute de Recherche Clinique du Bénin, Abomey-Calavi, Bénin
22 23 24	5 The DeWorm3 Project, University of Washington, Seattle, Washington, United States of America
24 25 26	6 Blantyre Institute for Community Outreach, Lions Sight First Eye Hospital, Blantyre, Malawi
26 27 28	7 Université de Paris, Institut de Recherche pour le Développement, MERIT, F-75006 Paris, France
28 29 30 21	8 Department of Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom
32 33 34	9 Department of Global Health, University of Washington, Seattle, Washington, United States of America
35 36 37	10 Department of Medicine (Infectious Diseases) and Pediatrics, University of Washington, Seattle, Washington, United States of America
38 38 30	11 Centre of Public Health Sciences, University of Iceland, Reykjavík, Iceland
40	*Corresponding Author Email: wsheahan@path.org

41 Abstract

42	Background: Soil Transmitted Helminths (STH) infect over 1.5 billion people globally and are associated
43	with anemia and stunting, resulting in an annual toll of 1.9 million Disability-Adjusted Life Years (DALYs).
44	School-based deworming (SBD), via mass drug administration (MDA) campaigns with albendazole or
45	mebendazole, has been recommended by the World Health Organization to reduce levels of morbidity
46	due to STH in endemic areas. DeWorm3 is a cluster-randomized trial, conducted in three study sites in
47	Benin, India, and Malawi, designed to assess the feasibility of interrupting STH transmission with
48	community-wide MDA as a potential strategy to replace SBD. This analysis examines data from the
49	DeWorm3 trial to quantify discrepancies between school-level reporting of SBD and gold standard
50	individual-level survey reporting of SBD.
51	
52	Methodology/Principal Findings: Population-weighted averages of school-level SBD calculated at the
53	cluster level were compared to aggregated individual-level SBD estimates to produce a Mean Squared
54	Error (MSE) estimate for each study site. In order to estimate individual-level SBD coverage, these MSE
55	values were applied to SBD estimates from the control arm of the DeWorm3 trial, where only school-
56	level reporting of SBD coverage had been collected.
57	
58	In each study site, SBD coverage in the school-level datasets was substantially higher than that obtained
59	from individual-level datasets, indicating possible overestimation of school-level SBD coverage. When
60	applying observed MSE to project expected coverages in the control arm, SBD coverage dropped from
61	89.1% to 70.5% (p-value < 0.001) in Benin, from 97.7% to 84.5% (p-value < 0.001) in India, and from
62	41.5% to 37.5% (p-value < 0.001) in Malawi.

64	Conclusions/Significance: These estimates indicate that school-level SBD reporting is likely to
65	significantly overestimate program coverage. These findings suggest that current SBD coverage
66	estimates derived from school-based program data may substantially overestimate true pediatric
67	deworming coverage within targeted communities.
68	
69	
70	
71	
72	
73	
74	
75	
/6 77	
// 70	
70 79	
80	
81	
82	
83	
84	
85	
86	
8/	
88 80	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99 100	
101	
102	
103	
104	
105	
106	

107 Introduction

108	Soil-transmitted helminths (STH) are a group of parasites (Ascaris lumbricoides, Ancylostoma duodenale,
109	<i>Necator americanus, and Trichuris trichiura</i>) estimated to infect over 1.5 billion people globally (<u>1</u>).
110	While these infections do not directly result in significant mortality, they do contribute substantially to
111	morbidity, causing malnutrition, anemia, and stunting, resulting in a toll of 1.9 million Disability-Adjusted
112	Life Years (DALYs) in 2019 ($\underline{2}$). To mitigate the impact of these parasitic infections, the World Health
113	Organization (WHO) has endorsed a strategy of deworming via mass drug administration (MDA) with
114	albendazole or mebendazole for pre-school-age children (PSAC) and school-age children (SAC), women
115	of childbearing age, and adults in high-risk occupations, including agricultural labor and mining.
116	Deworming campaigns in 2018 treated over 676 million SAC, representing approximately 53% of all
117	children estimated to be at risk for infection (1) .
118	
119	Measurement of program coverage is a critical aspect for deworming campaigns and is the primary
120	metric by which WHO sets programmatic targets. WHO guidance on conducting coverage surveys for
121	preventative chemotherapy defines coverage as answering the fundamental question "how many
122	people in need of treatment swallowed the drugs (<u>3</u>)." This guidance also notes that difficulty in
123	determining the underlying population denominator of SAC in a program area may impede the accurate
124	measurement of program coverage. School attendance may vary within program areas, and school
125	enrollment data may not accurately reflect the number of students that attend school on the day of
126	deworming treatment. Additionally, while many MDA campaigns for STH are conducted via school-
127	based drug distribution, school-based MDA may not reach all at-risk SAC, especially those who do not
128	attend school. To address this gap and protect against re-infection of SAC by disease reservoirs in the
129	broader community, a number of studies have examined the effect of conducting community-wide MDA
130	together with school-based MDA (<u>4</u>).

The DeWorm3 trial is a cluster-randomized trial examining the feasibility of interrupting the 131 132 transmission of STH by comparing community-wide MDA with the standard of care school-based MDA in 133 endemic areas of Benin, India, and Malawi (5). The analyses presented here use DeWorm3 trial data to 134 examine the degree to which discrepancies in school-level reporting of SBD may lead to overestimation 135 of SBD coverage when compared with individual-level reporting of SBD. Coverage estimates derived 136 from these two sources of data may differ, even in overlapping populations. School-level MDA data may 137 not reflect the same coverage levels as those from the individual-level MDA treatment records provided 138 by drug distribution data, as school-level coverage estimates might be based on a denominator that 139 includes only a subset of children who attend school on deworming days, potentially excluding SAC who 140 do not attend school on those days, or, as is the case in India, children aged 17 or 18 years who have 141 entered college and would be excluded from school-registry denominators. In Benin, children aged 5-14 142 years who were not enrolled in school were invited to go to their nearest school the day of MDA to be 143 treated, but only those enrolled in school were included in school-registry denominators. Previous 144 studies have suggested that such heterogeneous data reporting practices by schools during SBD may 145 result in overestimates of deworming coverage when compared to individual-level data collection (6). 146 147 DeWorm3 provides a novel opportunity to analyze these SAC coverage data due to the rigorous and 148 systematic conduct of annual household censuses that link the school each child attends to their 149 physical home address. This allows for an improved understanding of the geographic catchment areas of 150 each school in the study areas, as well as the underlying population denominators of each study cluster. 151 These data are less commonly collected by programs or by researchers, who often only collect school-152 level deworming data. As a result, this analysis offers an opportunity to assess overestimation in 153 traditional school-based coverage estimates. This comparison of school-level data to individual-level

- reporting has broad implications for understanding discrepancies in reported coverage for school-based
- 155 MDA programs.
- 156
- 157 Methods
- 158 <u>Study Design:</u> All data used in this analysis come from the DeWorm3 trial. A detailed description of the
- trial methods have been reported elsewhere (5). The analysis in this paper is descriptive in nature,
- 160 evaluating past results reported by the DeWorm3 trial by calculating new cluster-level coverage
- 161 estimates to compare with previously reported estimates.
- 162
- 163 Study Setting: The DeWorm3 trial operates in three primary locations. In Benin, the Institut de
- 164 Recherche Clinique du Bénin and the French Government's Institut de Recherche pour le
- 165 Développement collaborate with the Ministry of Health to administer the trial in the commune of Come.
- 166 In Malawi, trial operations are conducted by the Blantyre Institute for Community Outreach, the London
- 167 School of Hygiene and Tropical Medicine, and the Ministries of Health and Education in the district of
- 168 Mangochi. In India, trial operations are conducted in collaboration with the state government of Tamil
- 169 Nadu by the Christian Medical College, Vellore (5). Each country included an entire administrative area
- 170 (two administrative areas in India) inclusive of more than 80,000 individuals in each country site. These
- administrative areas were divided into 40 clusters, each containing a minimum of 1,650 individuals.
- 172 Twenty of these clusters were randomized to receive community-wide MDA, while twenty receive
- 173 standard of care SBD.

174

<u>Ethics statement</u>: The DeWorm3 trial has been reviewed and approved by the Institut de Recherche
Clinique au Bénin (IRCB) through the National Ethics Committee for Health Research (002-2017/CNERSMS) of the Ministry of Health in Benin, The London School of Hygiene and Tropical Medicine (12013),

178	The College of Medicine Research Ethics Committee (P.04/17/2161) in Malawi and the Christian Medical
179	College Institutional Review Board in Vellore, India (10392). The overall study was approved by The
180	Human Subjects Division at the University of Washington (STUDY00000180) and registered at
181	ClinicalTrials.gov (NCT03014167). This particular analysis was also reviewed and approved by the
182	Human Subjects Division at the University of Washington (Study 00012268).
183	
184	Consent procedures: Data collectors obtained consent / assent as appropriate prior to all data collection
185	activities. Written informed consent was sought if the individual could write, while oral consent was
186	given in the presence of a witness and documented with a thumbprint for any individual who could not
187	write.
188	
189	Study Population: As described by Ásbjörnsdóttir et al., 2018, the DeWorm3 trial tests the feasibility of
190	interrupting STH transmission throughout selected communities, using community-wide mass drug
191	distribution for all age groups in the intervention clusters. However, this analysis maintains as its focus
192	the treatment of school-age children (SAC) 5-14 years of age in Benin, 5-19 years of age in India, and 2-
193	19 years of age in Malawi, who are treated at school in both arms of the trial. A baseline census
194	conducted before the start of MDA in these three sites has been previously reported ($\overline{2}$) in which all
195	individuals from the households within each study area were enumerated and socio-demographic
196	characteristics collected.
197	
198	Data Collection: Data collected through the DeWorm3 trial were used for this analysis. Data collection
199	methods are detailed extensively in the original cluster-randomized trial protocol (6). The DeWorm3 trial
200	has provided door-to-door community-wide MDA for STHs in intervention arm clusters, timed to follow
201	the standard of care school-based MDA, while control clusters continue to receive standard of care

school-based MDA only. Community-wide coverage surveys assessing self-reported coverage and
treatment uptake in a subset of households in both arms are available. However, while individual-level
drug distribution data are available in the intervention arm, control arm data are limited to routine
school-level coverage data.

206

207 Data Analysis: Data were analyzed using R Studio 3.6.2 (8) and QGIS 3.10.7 A Coruña (9), primarily 208 leveraging the tidyverse package in R (10). Individual-level SBD coverage estimates were summarized to 209 the cluster level based on the location of each respondent's home. Because SAC are known to attend 210 schools in clusters that are different from the clusters where they reside, no single school-level coverage 211 estimate could be assumed to be directly representative of SBD coverage in the clusters in which the 212 schools were located. School-level coverage estimates were summarized to the cluster level using a 213 weighted average calculation. In this calculation, the reported coverage levels for all schools attended by 214 SAC in each cluster were weighted by multiplying them by the proportion of SAC in the cluster who 215 reported attending that school. For example, if there were 100 SAC in hypothetical cluster X, and thirty 216 of them self-reported as attending school A, fifty reported themselves attending school B, and twenty 217 reported themselves attending school C, the weights for schools A, B, and C, would be 0.3, 0.5, and 0.2 218 respectively.

219

This analysis leveraged the intervention clusters in which both individual-level and school-level SAC coverage estimates are available to quantify the discrepancy between these data sources. The individual-level data from community-wide MDA specifically identifies which SAC reported having been treated at school during the most recent round of school-based MDA. If we treat the individual-level estimates of previous coverage in each cluster as the reference value that acts as a gold-standard or "true" value, then over or under-estimation in the cluster-level coverage estimates from the school-level

- data can be measured by calculating the error, or the difference between those estimates and the gold-
- standard values from the individual-level data.
- 228
- 229 The difference between these "true" cluster-level estimates and the cluster-level estimates derived from
- the school-level data were calculated by subtracting each school-level estimate from each
- 231 corresponding individual-level estimate and squaring that value. These residual errors were then added
- together and divided by the number of clusters to get an estimate of the mean squared error (MSE) of
- the combined estimates for each study site. This statistic is available from the following equation:
- 234

235
$$MSE_{\hat{\theta}} = E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 = Var(\hat{\theta}) + (Bias of \hat{\theta})^2$$

236

237 Where $\hat{\theta}$ is the estimator of the unknown parameter θ (<u>11</u>), here taken to be the true SAC coverage at 238 the cluster-level which is approximated by individual-level coverage estimates from the intervention 239 arm.

240

241 After calculating the error in existing school-level SBD coverage estimates from the intervention arm, 242 the overall site errors were applied to control clusters where only school-level SBD estimates were 243 available. Multiplying the existing school-level SBD estimates by the ascertained error and subtracting 244 this from the school-level SBD estimate allowed us to project the likely individual-level SBD coverage 245 estimates in the control arm of the trial. Statistical significance for the difference in mean SBD coverage 246 estimates was produced by paired t-test (12) for Benin and Malawi where the cluster means were 247 normally distributed, and by Wilcoxon-Signed-Rank (13) test in India where the cluster means were 248 distributed significantly different from normal according to a Shapiro-Wilk normality test (14). 249

<u>Catchment Area Analysis:</u> The catchment areas of each school in the DeWorm3 registry were calculated
 to provide a more detailed understanding of where children within the study sites go for their SBD
 treatments. For each study site, the locations of every school in the trial's school registry were mapped
 in QGIS, along with the residential locations of each child for whom self-reported school attendance was
 available.

255

A key component of this analysis included the integration of free text fields from surveys that capture 256 257 school attendance for SAC treated in the community. While for the majority of SAC, their designated 258 school was chosen from dropdown menus and can be easily matched to the school they attend, many 259 respondents gave the names of schools that were not included in the trial's predetermined school 260 registries, or used alternative spellings, or colloquial names for their schools. These free-text responses 261 were matched to their official school names in the registries through a process of visual fuzzy matching 262 to identify common misspellings or translations, as well as quality control checks sent to study sites for 263 confirmation of assumptions on close spelling matches. Between 10 and 17% of free-text entries for 264 each site remained unmatched and were excluded from the catchment area creation process (Table 1).

265	Table 1. Proportion of School-Age Children in DeWorm3 Catchment Areas with Unmatched School
266	Names, by Study Site

	Matched (n)	Unmatched (n)	Total (n)	Proportion Unmatched
Benin	18140	3700	21840	0.17
India	21579	3929	25508	0.15
Malawi	38304	4246	42550	0.10

267

268 To obtain an initial understanding of the geographic range attributed to each school's catchment area,

the QGIS processing toolbox function "Concave Hull (k-nearest neighbor)" was utilized to create unique

270 minimum-bounding geographies that included the area within which SAC reported attending each

271 school. A naïve version of this analysis was conducted using all SAC whose responses were matched to a

- 272 specific school.
- 273
- 274 In order to obtain more discriminatory catchment areas for use in further analysis, the catchment area
- 275 creation process was repeated after removing statistical outliers by furthest geographical distance
- 276 between SAC and their self-reported schools. Using the QGIS processing toolbox function "Join by Lines
- 277 (hub lines)," the Euclidean distance was calculated between each SAC in the study area and the location
- of their self-reported school. In order to remove geographic outliers, the distribution of the distance
- variables was calculated separately for each study site, with histogram plots showing an approximately
- log-normal distribution in each case (Fig 1).

Fig 1. Histograms of the distance between School Age Children (SAC) and their self-reported schools.
 (A) Benin. (B) India. (C) Malawi. The left graph in each panel shows the distribution of the untransformed variable with distance measured in map units, on the right is the log-transformed variable.

284

285 Because of the approximately log-normal distribution of the distance between the SAC households and

- their reported schools, geographic outliers were removed for each school if the log-transformed
- distance between a respondent and their self-reported school was greater than the product of 1.96
- times the log of the standard deviation for each school plus the log of the mean distance between each
- school and the SAC who reported attending them.
- 290
- 291 Parent Study Power: No new sample size calculations were conducted for this analysis, as the requisite
- sample size powered to detect differences in final prevalence of STH infection between the control and
- 293 intervention arms was calculated by the DeWorm3 trial team during the course of previously published
- analyses (5). These calculations resulted in a minimum site sample size of 80,000 individuals, with
- individual clusters containing a minimum of 1,650 individuals.
- 296

297 Results

- 298 Cluster-level SAC SBD coverage was calculated using individual-level data from the clusters randomized
- to receive the community-wide MDA intervention annually in Malawi and Benin, and bi-annually in
- 300 India. These estimates specifically determined the cluster-level coverage as the proportion of SAC
- 301 reached during community-wide MDA who reported being treated during previous SBD. These SBD
- 302 coverages were consistently lower than the WHO 75% coverage goal for SAC and PSAC (<u>15</u>), with lower
- than 50% SBD coverage in 3 of 20 intervention clusters in India, 19 of 20 in Benin, and all 20 intervention
- 304 clusters in Malawi (Fig 2).

305 Fig 2. Individual-Level Reporting of SBD in Intervention Clusters, by Study Site.

306 (A) Benin. (B) India. (C) Malawi. Mean SAC Deworming Coverage shown as a red slicer line. Y-Axis set to
 307 WHO target of 75% coverage for SAC and PSAC.

- 308
- 309 These SAC SBD coverages from individual-level data in the intervention clusters were compared to
- school-level data collected through the standard of care SBD that occurred in all clusters. By comparing
- 311 the school-level data in the intervention clusters to the gold standard individual-level data in those same
- 312 clusters, we calculated mean squared errors (MSE) for each cluster which were then averaged to
- determine a single error value for each site. These errors were 0.096 in Malawi, 0.135 in India, and 0.209
- 314 in Benin (<u>Fig 3</u>).

315 Fig 3. Comparison of cluster-level SAC SBD coverage from individual-level data collection to school-

- 316 level SBD in intervention clusters, by site.
- 317 (A) Benin. (B) India. (C) Malawi.
- 318
- 319 The overall MSE measurements per site were applied to school-level SBD coverages in the control
- 320 clusters to project individual-level SBD coverage estimates for the control arm of the DeWorm3 trial (Fig
- 321 <u>4)</u>.

Fig 4. Comparison of school-level SBD coverage estimates to projected SBD coverage estimates in control clusters if individual-level data had been available, by study site.

- 324 (A) Benin. (B) India. (C) Malawi.
- 325

326	In Ronin		data from	the trial's	control clusters	originally	showed an	JVAR200 SAC
520	III Dellill,	SCHOOL-IEVEL 3DD	uata mom	the that s	control clusters	Uligiliali	y showed an	average SAC

- 327 coverage of 89.1%, compared to 70.5% (p-value < 0.001) when applying observed MSE to project
- 328 expected coverages if individual-level data had been available. In India, 97.7% SAC school-level SBD
- 329 coverage for control clusters decreased to 84.5% (p-value < 0.001) when projected for SBD estimates
- 330 from individual-level data, and in Malawi the SAC school-level SBD coverage decreased from 41.5% to
- 331 37.5% (p-value < 0.001) when projected for SBD estimates from individual-level data (Fig 5).

332 Fig 5. Comparison of school-level average SBD coverage to average individual-level SBD coverage in 333 control clusters after application of mean-squared error rates.

- 334
- 335 The catchment areas created in this analysis show a high degree of overlap, particularly in the highly
- 336 populated urban centers of the DeWorm3 study areas. Removal of geographic outliers resulted in fewer
- instances where catchment areas were greatly extended in order to include a small number of distant 337
- 338 SAC. Recalculated school catchment areas were shown to overlap each other less than in the naïve
- 339 analysis, particularly in Benin (Fig 6A).

340 Fig 6. Catchment areas for each school by study area, created via QGIS Concave Hull (K-nearest 341 neighbor) analysis. 342 (A) Benin. (B) India. (C) Malawi. The left side of each panel shows naïve catchment areas in semi-

343 transparent pink created prior to removal of geographic outliers, while the right side shows catchment

areas in semi-transparent green created after removal of geographic outliers. Both overlay original 344 345 cluster boundaries in green.

346

347 This reflects the reality that school catchment areas are geographically indistinct and do not merely

- 348 include the SAC closest to them. This means that any attempt to fix catchment areas solely using
- 349 geographic proximity is unlikely to be accurate, especially in densely populated urban centers with many
- 350 schools. For example, of the 40 clusters comprising the Benin study area, in only 7 clusters did more
- 351 than 50% of SAC attend the school that was closest to their home by Euclidean distance, and in only one
- 352 cluster was this the case for more than 65% of SAC (Supplemental Table 1). Overall, the percentage of

- 353 SAC that attended the school geographically closest to their home was 29% in Benin and 10% in India,
- although that number was substantially higher in Malawi at 66% (Table 2).

Table 2. Number of School-Age Children in DeWorm3 Catchment Areas and Proportion Attending the School Geographically Nearest to Their Household, by Study Site

	· · · ·	
	Number of SAC (n)	Proportion Attending Nearest School
Benin	21840	0.29
India	25508	0.11
Malawi	42550	0.66

357

358 Discussion

359 The success of global STH programs is critically dependent on the accurate measurement of

360 programmatic treatment coverage. Funding mechanisms, policy directives, and the support of local

361 stakeholders all require that programs can demonstrate the reach and the impact that they provide to

362 local communities. Accurate reporting of data is of fundamental importance to achieving this goal. This

363 paper seeks to enhance our understanding of deworming coverage provided during SBD by

364 demonstrating the error inherent in school-level data collection when compared to a gold standard of

365 individual-level reporting. This paper compares individual-level SBD data collected via community-wide

366 MDA from the DeWorm3 trial's intervention arm against school-level SBD data collected via school-level

367 reporting in both arms to calculate the difference between these estimates in the intervention arm

368 clusters where both numbers are available. This is a new approach for characterizing discrepancies of

369 SAC deworming coverage during SBD. This analysis was made possible by the DeWorm3 trial's recording

of previous SBD treatment status for SAC reached during community-level MDA.

371

Cluster-level SBD coverages for SAC derived from individual-level data were substantially lower than SBD
 estimates derived from school-level reporting, which indicates that school-level SBD reporting may be

374 over-estimating SAC deworming coverage when compared to a gold standard of individual-level 375 reporting. This is an important finding because school-level estimates for SAC deworming are more commonly available to national STH deworming programs than the individual-level data provided in a 376 377 trial such as DeWorm3. If school-level data are consistently overestimating coverage, then decisions 378 made using these data at the program level may be made under the assumption that SBD coverage is 379 higher than it actually is. Such a discrepancy could result in miscalculation of the duration and frequency 380 of deworming for future planning, premature cessation of MDA or disruptive allocation of resources 381 based on a flawed assumption of treatment coverage.

382

383 There may be a number of reasons for the differences between these estimates that are not accounted 384 for by systematic over-estimation of school-level SBD coverage. It is important to note that the weighted 385 average calculation used to determine school-level SBD coverages in this analysis is most heavily 386 influenced by the schools that report the largest number of SAC attendees from school enrollment data 387 in the DeWorm3 trial's annual school census. If a small number of SAC from a particular cluster report 388 attending a school with low coverage, then that cluster may still have a large number of SAC who attend 389 a nearby school with higher coverage which would decrease the low-coverage signal. This would lead to 390 a possible overestimation of cluster-level coverage when relying solely upon school-level data if higher 391 coverages are associated with larger schools. Additionally, the school-enrollment totals available during 392 SBD may have included SAC that do not usually attend that school but who showed up for SBD days. This 393 may have been the case in Benin and India, where overall enrollment totals on the day of SBD exceeded 394 enrollment totals in the trial's school census, as opposed to in Malawi, where there were fewer students 395 listed as enrolled on the day of deworming than in the school census. If these SAC who are arriving for 396 deworming treatment days but are not regularly enrolled are more difficult to reach during community-397 wide MDA data collection, then individual-level reporting in the intervention arm may be

underestimating true treatment coverage of SBD, leading to an erroneously high depiction of overestimation in the school-level SBD data.

400

401 The finding that coverage reported by schools or district health authorities is consistently higher than 402 coverage as determined by individual-level survey responses has been previously observed in trials (16, 403 17, 18), and when comparing national reporting of school-based treatment to WHO versus individual-404 level Demographic and Health Survey results (19). Previous papers have postulated that the use of 405 school or district-level reported coverages may open studies to biased results that overestimate 406 deworming coverage due to exaggeration by community drug distributors, local health officials, or 407 schoolteachers (6). Another standard concern for survey-based data collection is recall bias, although 408 this is unlikely to have affected the low coverages reported during individual-level data collection, as the 409 DeWorm3 trial utilized indelible ink to mark the fingers of SAC who were treated at school in order to 410 mitigate this concern (20). In discussions with DeWorm3 program staff, it was noted that SBD coverages 411 derived from individual-level data collection for the data analyzed in this paper might have been 412 unusually low in Malawi because of a longer-than-usual gap between SBD and community-wide drug 413 distribution during that year, which would normally occur within two weeks. It is possible that this may have resulted in lower estimates for previous treatment during SBD because the ink may have faded 414 415 from the fingers of SAC, preventing the use of this marking for accurate measurement of previous SBD 416 treatment.

417

This paper also utilized concave-hull geospatial analysis to define the catchment areas of schools in the DeWorm3 trial areas for the purposes of estimating SAC deworming treatment coverage. Previous approaches have relied upon proximity-based (<u>21</u>), Bayesian modeling (<u>22</u>), and cost-distance approaches (<u>23</u>) to estimate likely catchment areas for service delivery of health posts, hospitals, and

422 schools for deworming activities. The DeWorm3 trial provides access to geolocation information for 423 each SAC treated within the study areas, allowing for exact visualization of reported catchment areas. 424 The school catchment areas in the DeWorm3 trial areas overlap each other significantly, especially in the 425 population centers where schools are geographically near to one another. In addition, most SAC in these 426 study areas did not receive their deworming treatments at the school geographically nearest to them. If 427 this is the case for populations in other study areas, this means that proximity-based approaches for 428 defining health activity catchment areas may introduce considerable inaccuracies for determining 429 relevant population denominators. A lack of clear communication to survey administrators regarding the 430 geographic limits of clusters, municipal boundaries, and catchment areas has been previously identified 431 as a possible driver of variation in deworming treatment coverage (24, 6). Therefore, a potential benefit 432 stemming from the visualization of these catchment areas may include more streamlined data collection 433 and drug-distribution practices for school-based and community-wide MDA in the DeWorm3 trial area. 434 435 The DeWorm3 trial is an exceptionally large and rigorously conducted study, with harmonized data 436 collection practices across three different countries. Additionally, these data had already gone through

437 initial data cleaning from the study sites and the central DeWorm3 data team, resulting in overall
438 uniformity and confidence for use in analysis.

439

However, there were a number of limitations that should be considered when interpreting these
findings. The uneven quality of self-reported school attendance data used to create catchment areas
and to assign population-based weights to each school in the cluster-level weighted average coverage
calculations may have impacted these results. In the case of catchment-area creation, naïve catchments
from the self-reported data were modified to exclude wide-ranging geographic outliers to more
accurately predict where a hypothetical child for whom self-reported school attendance is unavailable

446 would likely go for deworming treatments in a given cluster. And for the assignment of population-

- 447 based weights, the self-reported attendance data do not allow for consideration of those SAC who may
- 448 not have been reached during data collection, or whose responses were matched to the incorrect
- school.
- 450

451 Free-text school name entries also pose a limitation because there is no perfect recording of where each 452 child in a given cluster received their SBD treatment. To minimize the effect of this limitation, thousands 453 of free-text entries were analyzed and matched, where possible, to the appropriate school in each study 454 site's school registry, via the iterative process described in the methods section. In this analysis, 17% of 455 SAC respondents in Benin were not matched to a specific school, while 15% and 10% of SAC remained 456 unmatched to a school in India and Malawi, respectively. These unmatched SAC reported going to 457 schools that were unknown, were outside the study area, or were schools for which coverage levels and 458 enrollment data were not available, with informal or nursery schools accounting for the largest 459 percentage of free-text entries that cannot be matched to a registered school. As an illustrative 460 example, it is believed that over 97% of the 3,929 unmatched free text entries for the India study site of 461 the DeWorm3 trial refer to Anganwadi (informal government-run education centers for rural childcare 462 developed under the Integrated Child Development Scheme) centers due to similarities in the submitted 463 free-text responses. This has been a limitation in previous studies, including in Bangladesh, where 464 informal schooling complicated calculation of coverage and population statistics (25). If large numbers of these unmatched respondents actually attended one of the schools in the school registry, then the 465 466 weighted average formula for determining coverage in SBD from school-level data would have slightly 467 underweighted the schools that they attended. This would only provide a significant limitation if large 468 numbers of SAC were assumed to attend an erroneous school in the registry, leading to a slight 469 overweighting of that school in the weighted average coverage formula.

., .

471	Finally, it is important to note that this analysis only utilized data collected in year two of the DeWorm3
472	trial, which consisted of two treatment rounds in India and 1-2 rounds in Benin and Malawi, as these
473	data were determined by the DeWorm3 program staff to be the most complete dataset available. In
474	year one, data were less reliable because certain data collection processes had not yet been
475	standardized, and in year three, data collection was disrupted by the worldwide emergence of SARS-
476	CoV-2. A more complete analysis would examine data across multiple study years to determine if SAC
477	coverage rates fluctuated significantly from year to year, and if observed error in the school-level SBD
478	data were sustained throughout the duration of the trial.
479	
480	Conclusion
481	This analysis utilized data from the DeWorm3 trial to quantify discrepancies between cluster-level
482	estimates of SAC deworming coverage derived from gold-standard individual-level data collection during
483	community-wide MDA in the intervention arm and historically less reliable school-level data collection
484	during the delivery of standard of care school-based deworming in both trial arms. These estimates,
485	derived from three different country sites and across many settings within those countries, indicate that
486	school-level reporting of SAC deworming coverage likely overestimate program coverage in these age
487	groups. These results suggest that school-level reporting of SAC deworming coverage should be
488	validated by individual-level data collection. Novel interventions to improve data collection within MDA
489	programs are needed to ensure that accurate reporting informs programmatic decision-making and
490	resource allocation.
491 492 493 494 495	

496 Acknowledgements

The authors wish to thank all of the study participants, communities and community leaders,
national NTD program staff and other local, regional and national partners who have participated or
supported the DeWorm3 study. We also would like to acknowledge the work of all members of the
DeWorm3 study teams and affiliated institutions. We give special thanks to the Clinical Trial and
Implementation Science Support Unit at the University of Washington for their aid in the conception
of this analysis.

503

506

507

508

- 504 505 **References**
 - Soil-transmitted helminth infections [Internet]. World Health Organization. 2020 [cited 2021 Jun 10]. Available from: <u>https://www.who.int/news-room/fact-sheets/detail/soil-transmittedhelminth-infections</u>
- Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369
 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the
 Global Burden of Disease Study 2019. The Lancet. 2020 Oct 17;396(10258):1204–22.
- 5123. Coverage Evaluation Guidelines Final Draft_Nov 2016.pdf [Internet]. [cited 2021 Jun 10].513Available from:
- 514https://www.ntdsupport.org/sites/default/files/uploads/docs/resources/Coverage%20Evaluatio515n%20Guidelines%20Final%20Draft_Nov%202016.pdf
- Burnim M, Ivy JA, King CH. Systematic review of community-based, school-based, and combined delivery modes for reaching school-aged children in mass drug administration programs for schistosomiasis. PLOS Neglected Tropical Diseases. 2017 Oct 27;11(10):e0006043.
- 5. Ásbjörnsdóttir KH, Ajjampur SSR, Anderson RM, Bailey R, Gardiner I, Halliday KE, et al. Assessing the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration: The DeWorm3 cluster randomized trial protocol. PLOS Neglected Tropical Diseases. 2018 Jan 18;12(1):e0006166.
- Onkanga IO, Mwinzi PNM, Muchiri G, Andiego K, Omedo M, Karanja DMS, et al. Impact of two
 rounds of praziquantel mass drug administration on Schistosoma mansoni infection prevalence
 and intensity: a comparison between community wide treatment and school based treatment in
 western Kenya. International Journal for Parasitology. 2016 Jun 1;46(7):439–45.
- Team TDT. Baseline patterns of infection in regions of Benin, Malawi and India seeking to
 interrupt transmission of soil transmitted helminths (STH) in the DeWorm3 trial. PLOS Neglected
 Tropical Diseases. 2020 Nov 2;14(11):e0008771.
- 5308. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL531http://www.rstudio.com/.
- 9. QGIS Development Team, 2021. QGIS Geographic Information System. Open Source Geospatial
 Foundation. URL <u>http://qgis.org</u>
- 534 10. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43),
 535 1686, https://doi.org/10.21105/joss.01686
- 53611. Zheng Methods of Evaluating Estimators.pdf [Internet]. [cited 2021 Jun 10]. Available from:537http://people.missouristate.edu/songfengzheng/teaching/mth541/lecture%20notes/evaluation.538pdf

12. Ross A., Willson V.L. (2017) Paired Samples T-Test. In: Basic and Advanced Statistical Tests. SensePublishers, Rotterdam. <u>https://doi-org.offcampus.lib.washington.edu/10.1007/978-94-</u> 6351-086-8_4

542	13.	Rey D., Neuhäuser M. (2011) Wilcoxon-Signed-Rank Test. In: Lovric M. (eds) International
543		Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. <u>https://doi-</u>
544		org.offcampus.lib.washington.edu/10.1007/978-3-642-04898-2_616
545	14.	Shapiro, S.S. & Wilk, M.B. (1965) An analysis of variance for normality (complete samples).
546		Biometrika, Vol. 52, No. 3/4.
547	15.	World Health Organization, Nutrition for Health and Development, World Health Organization,
548		Department of Control of Neglected Tropical Diseases, World Health Organization. Preventive
549		chemotherapy to control soil-transmitted helminth infections in at-risk population groups:
550		guideline. [Internet]. 2017 [cited 2021 Jun 10]. Available from:
551		http://www.ncbi.nlm.nih.gov/books/NBK487927/
552	16.	Katabarwa MN, Griswold E, Habomugisha P, Eyamba A, Byamukama E, Nwane P, et al.
553		Comparison of Reported and Survey-Based Coverage in Onchocerciasis Programs over a Period
554		of 8 Years in Cameroon and Uganda. The American Journal of Tropical Medicine and Hygiene.
555		2019 Mar 25;100(5):1208–15.
556	17.	Worrell C, Mathieu E. Drug coverage surveys for neglected tropical diseases: 10 years of field
557		experience. Am J Trop Med Hyg. 2012 Aug;87(2):216–22.
558	18.	Binder S, Campbell CH, Castleman JD, Kittur N, Kinung'hi SM, Olsen A, et al. Lessons
559		Learned in Conducting Mass Drug Administration for Schistosomiasis Control and
560		Measuring Coverage in an Operational Research Setting, Am I Trop Med Hyg. 2020
561		lul·103(1 Suppl)·105–13
562	19	Lo NC Gunta B. Addiss DG. Bendavid E. Heft-Neal S. Mikhailov A. et al. Comparison of World
563	19.	Health Organization and Demographic and Health Surveys data to estimate sub-national
564		deworming coverage in pre-school aged children. PLOS Neglected Tronical Diseases 2020 Aug
565		17·14/8)·e0008551
566	20	Cromwell FA Ngondi L McFarland D King ID Emerson PM Methods for estimating
500	20.	nonulation coverage of mass distribution programmes: a review of practices in relation
507		to trachering control. Transactions of The Devel Conjety of Transact Medicine and
508		to trachoma control. Transactions of The Royal Society of Tropical Medicine and
569	~ 4	Hyglene. 2012 Oct 1;106(10):588–95.
570	21.	Cattaneo A, Nelson A, McMenomy I. Global mapping of urban–rural catchment areas reveals
571		unequal access to services. PNAS [Internet]. 2021 Jan 12 [cited 2021 Jun 10];118(2). Available
572	~ ~	from: http://www.pnas.org/content/118/2/e2011990118
5/3	22.	Alegana VA, Khazenzi C, Akech SO, Snow RW. Estimating hospital catchments from in-patient
574		admission records: a spatial statistical approach applied to malaria. Sci Rep. 2020 Jan
575	•••	
576	23.	Stresman GH, Stevenson JC, Owaga C, Marube E, Anyango C, Drakeley C, et al. Validation of
577		three geolocation strategies for health-facility attendees for research and public health
578		surveillance in a rural setting in western Kenya. Epidemiology & Infection. 2014
579	~ 4	Sep;142(9):1978–89.
580	24.	Pullan RL, Halliday KE, Oswald WE, Micharo C, Beaumont E, Kepha S, et al. Effects, equity, and
581		cost of school-based and community-wide treatment strategies for soil-transmitted neiminths in
582	25	Kenya: a cluster-randomised controlled trial. The Lancet. 2019 May 18;393(10185):2039–50.
583	25.	Nath TC, Padmawati RS, Murhandarwati EH. Barriers and gaps in utilization and coverage of
584		mass drug administration program against soil-transmitted neiminth infection in Bangladesh: An
585		implementation research. Journal of infection and Public Health. 2019 Mar 1;12(2):205–12.
586		
587		

588 Supporting Information

- 589 S1 Table. Cluster-level School Age Children Statistics, by Study Site Cluster-level statistics of the
- 590 proportion of SAC attending the school that is geographically closest to their home, the total number of
- 591 students, and the number of unique schools reported as being attended by students in each cluster,
- replicated for each of the three study sites.













School-Level Data

Projected Individual-Level Data









