PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Overestimation of school-based deworming coverage resulting from school-based reporting

William Sheahan[1]*, Roy Anderson[2], Kumudha Aruldas[3], Euripide Avokpaho[4], Sean Galagan[5], Jeanne Goodman[5], Parfait Houngbegnon[4], Gideon John Israel[3], Venkateshprabhu Janagaraj[3], Saravanakumar Puthupalayam Kaliappan[3], Arianna Rubin Means[5], Chloe Morozoff[5], Emily Pearman[5], Rohan Michael Ramesh[3], Amy Roll[5], Alexandra Schaefer[5], James Simwanza[6], Stefan Witek-McManus[7], Sitara S. R. Ajjampur[3‡], Robin Bailey[6‡], Moudachirou Ibikounlé[4‡], Khumbo Kalua[6‡], Adrian J. F. Luty[8‡], Rachel Pullan[7‡], Judd L. Walson[5,9,10☯], Kristjana Hrönn Ásbjörnsdóttir[5,11☯]

1 Malaria and Neglected Tropical Diseases, PATH, Seattle, Washington, United States of America, **2** School of Public Health, Faculty of Medicine, Imperial College, London, United Kingdom, **3** The Wellcome Trust Research Laboratory, Division of Gastrointestinal Sciences, Christian Medical College, Vellore, India, **4** Institute de Recherche Clinique du Bénin, Abomey-Calavi, Bénin, **5** The DeWorm3 Project, University of Washington, Seattle, Washington, United States of America, **6** Blantyre Institute for Community Outreach, Lions Sight First Eye Hospital, Blantyre, Malawi, **7** Department of Disease Control, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom, **8** Université de Paris, Institut de Recherche pour le Développement, MERIT, Paris, France, **9** Department of Global Health, University of Washington, Seattle, Washington, United States of America, **10** Department of Medicine (Infectious Diseases) and Pediatrics, University of Washington, Seattle, Washington, United States of America, **11** Centre of Public Health Sciences, University of Iceland, Reykjavík, Iceland

☯ These authors contributed equally to this work.
‡ SSRA, RB, MI, KK, AJFL and RP also contributed equally to this work.
* wsheahan@path.org

OPEN ACCESS

Citation: Sheahan W, Anderson R, Aruldas K, Avokpaho E, Galagan S, Goodman J, et al. (2023) Overestimation of school-based deworming coverage resulting from school-based reporting. PLoS Negl Trop Dis 17(4): e0010401. https://doi.org/10.1371/journal.pntd.0010401

Editor: Xiao-Nong Zhou, NIPD: National Institute of Parasitic Diseases, CHINA

Received: April 8, 2022

Accepted: November 21, 2022

Published: April 10, 2023

Data Availability Statement: Data cannot be shared publicly because the study remains blinded to outcome data. Data are available from the DeWorm3 Institutional Data Access Committee (contact via deworm3@uw.edu, https://depts.washington.edu/deworm3/) for researchers who meet the criteria for access to these data.

Funding: This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [Grant Number INV-022149 (received by JLW)]. Under the grant conditions of the Foundation, a Creative

Abstract

Background

Soil Transmitted Helminths (STH) infect over 1.5 billion people globally and are associated with anemia and stunting, resulting in an annual toll of 1.9 million Disability-Adjusted Life Years (DALYs). School-based deworming (SBD), via mass drug administration (MDA) campaigns with albendazole or mebendazole, has been recommended by the World Health Organization to reduce levels of morbidity due to STH in endemic areas. DeWorm3 is a cluster-randomized trial, conducted in three study sites in Benin, India, and Malawi, designed to assess the feasibility of interrupting STH transmission with community-wide MDA as a potential strategy to replace SBD. This analysis examines data from the DeWorm3 trial to quantify discrepancies between school-level reporting of SBD and gold standard individual-level survey reporting of SBD.

Methodology/Principal findings

Population-weighted averages of school-level SBD calculated at the cluster level were compared to aggregated individual-level SBD estimates to produce a Mean Squared Error (MSE) estimate for each study site. In order to estimate individual-level SBD coverage,

Competing interests: The authors have declared that no competing interests exist.

these MSE values were applied to SBD estimates from the control arm of the DeWorm3 trial, where only school-level reporting of SBD coverage had been collected.

In each study site, SBD coverage in the school-level datasets was substantially higher than that obtained from individual-level datasets, indicating possible overestimation of school-level SBD coverage. When applying observed MSE to project expected coverages in the control arm, SBD coverage dropped from 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 41.5% to 37.5% (p-value < 0.001) in Malawi.

Conclusions/Significance

These estimates indicate that school-level SBD reporting is likely to significantly overestimate program coverage. These findings suggest that current SBD coverage estimates derived from school-based program data may substantially overestimate true pediatric deworming coverage within targeted communities.

Trial registration

NCT03014167.

Author summary

Soil Transmitted Helminths (STH) infect over 1.5 billion people globally and are associated with anemia and stunting. School-based deworming (SBD) via mass drug administration has been recommended by the World Health Organization to reduce levels of morbidity due to STH in endemic areas. This study examines discrepancies between school-level reporting of SBD and individual-level survey reporting of SBD from the DeWorm3 trial in Benin, India, and Malawi. Study cluster coverages were calculated for both school-level and individual-level data in the intervention arm and the differences between them were used to produce a mean squared error (MSE) statistic for each cluster. These values were then applied to the control arm, where only school-level data had been collected, to estimate individual-level data. In each study site, SBD coverage in the school-level data was substantially higher than in the individual-level data. When applying observed MSE to project individual-level data in the control arm, SBD coverage dropped from 89.1% to 70.5% in Benin, from 97.7% to 84.5% in India, and from 41.5% to 37.5% in Malawi. These estimates indicate that current SBD coverage estimates derived from school-level data may substantially overestimate true pediatric deworming coverage within targeted communities.

Introduction

Soil-transmitted helminths (STH) are a group of parasites (*Ascaris lumbricoides*, *Ancylostoma duodenale*, *Necator americanus*, *and Trichuris trichiura*) estimated to infect over 1.5 billion people globally [1]. While these infections do not directly result in significant mortality, they do contribute substantially to morbidity, causing malnutrition, anemia, and stunting, resulting in a toll of 1.9 million Disability-Adjusted Life Years (DALYs) in 2019 [2]. To mitigate the impact of these parasitic infections, the World Health Organization (WHO) has endorsed a

strategy of deworming via mass drug administration (MDA) with albendazole or mebendazole for pre-school-age children (PSAC) and school-age children (SAC), women of childbearing age, and adults in high-risk occupations, including agricultural labor and mining. Deworming campaigns in 2018 treated over 676 million SAC, representing approximately 53% of all children estimated to be at risk for infection [1].

Measurement of program coverage is a critical aspect for deworming campaigns and is the primary metric by which WHO sets programmatic targets. WHO guidance on conducting coverage surveys for preventative chemotherapy defines coverage as answering the fundamental question "how many people in need of treatment swallowed the drugs [3]." This guidance also notes that difficulty in determining the underlying population denominator of SAC in a program area may impede the accurate measurement of program coverage. School attendance may vary within program areas, and school enrollment data may not accurately reflect the number of students that attend school on the day of deworming treatment, or the precise number of students who ultimately decide to accept the drug, taking into account refusal of treatment due to concerns over side effects, or on the basis of religious, traditional, or other beliefs [4]. Additionally, while many MDA campaigns for STH are conducted via school-based drug distribution, school-based MDA may not reach all at-risk SAC, especially those who do not attend school. To address this gap and protect against re-infection of SAC by disease reservoirs in the broader community, a number of studies have examined the effect of conducting community-wide MDA together with school-based MDA [5–8].

The DeWorm3 trial is a cluster-randomized trial examining the feasibility of interrupting STH transmission throughout selected communities, using community-wide mass drug distribution for all age groups in the intervention clusters [9]; however this analysis focuses on the treatment of school-age children (SAC). In both arms of the trial, SAC 5–14 years of age in Benin, 5–19 years of age in India, and 2–19 years of age in Malawi, were treated at school according to national guidelines (twice annually in India, and once annually in Benin and Malawi). Community-wide MDA was planned to follow school-based deworming, and during drug distribution visits children in the intervention clusters who were treated at school were identified by finger inking and/or self-report and their treatment history was recorded, while those who were missed by school-based deworming had the opportunity to be treated along with other household members. The baseline prevalence of any STH infection was highest at the India site (17.0%), followed by Malawi (7.4%) and Benin (5.3%) [10].

The analyses presented here use DeWorm3 trial data to examine the degree to which discrepancies in school-level reporting of SBD may lead to overestimation of SBD coverage when compared with individual-level reporting of SBD. Coverage estimates derived from these two sources of data may differ, even in overlapping populations. School-level MDA data may not reflect the same coverage levels as those from the individual-level MDA treatment records provided by drug distribution data, as school-level coverage estimates might be based on a denominator that includes only a subset of children who attend school on deworming days, potentially excluding SAC who do not attend school on those days, or, as is the case in India, children aged 17 or 18 years who have entered college and would be excluded from school-registry denominators. In Benin, children aged 5–14 years who were not enrolled in school were invited to go to their nearest school the day of MDA to be treated, but only those enrolled in school were included in school-registry denominators. At the Malawi site, school-based deworming was carried out by the DeWorm3 team and included praziquantel (due to endemicity of schistosomiasis) in addition to albendazole. SAC not enrolled in school, as well as PSAC aged 2–4 and older children aged 15–19, were eligible to participate in the school-based deworming, despite not being listed in school-registry denominators. Previous studies have suggested that such heterogeneous data reporting practices by schools during SBD may

result in overestimates of deworming coverage when compared to individual-level data collection [11].

DeWorm3 provides a novel opportunity to analyze these SAC coverage data due to the rigorous and systematic conduct of annual household censuses that link the school each child attends to their physical home address. This allows for an improved understanding of the geographic catchment areas of each school in the study areas, as well as the underlying population denominators of each study cluster. These data are less commonly collected by programs or by researchers, who often only collect school-level deworming data. As a result, this analysis offers an opportunity to assess overestimation in 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 MDA programs.

Methods

Ethics statement

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/CNERS-MS) of the Ministry of Health in Benin, The London School of Hygiene and Tropical Medicine (12013), The College of Medicine Research Ethics Committee (P.04/17/2161) in Malawi and the Christian Medical College Institutional Review Board in Vellore, India (10392). The DeWorm3 trial was also approved by The Human Subjects Division at the University of Washington (STUDY00000180) and registered at ClinicalTrials.gov (NCT03014167). This particular analysis was reviewed and approved by the Human Subjects Division at the University of Washington (Study 00012268).

Consent procedures

Data collectors obtained consent/assent as appropriate prior to all data collection activities. Informed consent for each household's participation in the Deworm3 trial, including the census and drug administration activities, was sought from the head of household. Written consent was sought if the individual could write, while oral consent was given in the presence of a witness and documented with a thumbprint for any individual who could not write. The data in this manuscript were collected as part of household drug distribution activities and were determined not to require consent or assent beyond the household's consent for participation in the trial.

Study design

All data used in this analysis come from the DeWorm3 trial. A detailed description of the trial methods have been reported elsewhere [9]. The analysis in this paper is descriptive in nature, evaluating past results reported by the DeWorm3 trial by calculating new cluster-level coverage estimates to compare with previously reported estimates.

The DeWorm3 Trial study sites were selected by an independent committee that sought to include a broad range of geographic and epidemiologic settings through a rigorous process previously detailed elsewhere [10], with the primary evaluation criteria consisting of baseline prevalence estimates, a history of MDA for lymphatic filariasis, and existing support for participation from local officials as well as national STH program managers. Each country site included an area that was inclusive of more than 80,000 individuals. These areas were divided into 40 clusters, each containing a minimum of 1,650 individuals. Twenty of these clusters

were randomized to receive community-wide MDA, while twenty received standard of care SBD.

Study setting

The DeWorm3 trial operates in three primary locations. In Benin, the Institut de Recherche Clinique du Bénin and the French Government's Institut de Recherche pour le Développement collaborate with the Ministry of Health to administer the trial in the commune of Come. In Malawi, trial operations are conducted by the Blantyre Institute for Community Outreach, the London School of Hygiene and Tropical Medicine, and the Ministries of Health and Education in the district of Mangochi. In India, trial operations are conducted in collaboration with the state government of Tamil Nadu by the Christian Medical College, Vellore [9].

Study population

As described by Ásbjörnsdóttir et al., 2018, [9] the DeWorm3 trial tests the feasibility of interrupting STH transmission throughout selected communities, using community-wide mass drug distribution for all age groups in the intervention clusters. However, this analysis maintains as its focus the treatment of school-age children (SAC) 5–14 years of age in Benin, 5–19 years of age in India, and 2–19 years of age in Malawi, who are treated at school in both arms of the trial. A baseline census conducted before the start of MDA in these three sites has been previously reported [12] in which all individuals from the households within each study area were enumerated and socio-demographic characteristics collected.

Data collection

Data collected through the DeWorm3 trial were used for this analysis. Data collection methods are detailed extensively in the original cluster-randomized trial protocol [9]. In brief, the DeWorm3 trial has provided door-to-door community-wide MDA for STHs in intervention arm clusters, timed to follow 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.

Outcome definitions

Individual-level Cluster SAC SBD coverages were calculated using data collected from all resident children in the clusters randomized to receive the community-wide MDA intervention. During community-wide MDA, children were asked whether they had been treated during the recent SBD, and their fingernails were checked for ink applied during SBD for participating children. Individual-level cluster SBD coverage was defined as the proportion of children in the most recent cluster census who were determined to have been treated during SBD by these methods.

School-level SAC SBD coverage estimates were derived from school-based reporting of the numbers of children treated at each school on deworming days, divided by the number of children reported as attending that school. As school catchment areas did not directly correspond to trial clusters, cluster-level coverage estimates were derived from school-based reporting as detailed below.

Data analysis

Data were analyzed using R Studio 3.6.2 [13] and QGIS 3.10.7 A Coruña [14], primarily leveraging the tidyverse package in R [15]. Individual-level SBD coverage estimates were summarized to the cluster level based on the location of each respondent's home. Because SAC throughout the study areas are frequently known to attend schools in clusters that are different from the clusters where they reside, no single school-level coverage estimate could be assumed to be directly representative of SBD coverage in the clusters in which the schools were located. School-level coverage estimates were summarized to the cluster level using a weighted average calculation. In this calculation, the reported coverage levels for all schools attended by SAC in each cluster were weighted by multiplying them by the proportion of SAC in the cluster who reported attending that school. For example, if there were 100 SAC in hypothetical cluster X, and thirty of them self-reported as attending school A, fifty reported themselves attending school B, and twenty reported themselves attending school C, the weights for schools A, B, and C, would be 0.3, 0.5, and 0.2 respectively.

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.

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 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:

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

Where $\hat{\theta}$ is the estimator of the unknown parameter θ [16], here taken to be the true SAC coverage at the cluster-level which is approximated by individual-level coverage estimates from the intervention arm.

After calculating the error in existing school-level SBD coverage estimates from the intervention arm, the overall site errors were applied to control clusters where only school-level SBD estimates were available. Multiplying the existing school-level SBD estimates by the ascertained error and subtracting this from the school-level SBD estimate allowed us to project the likely individual-level SBD coverage estimates in the control arm of the trial. Statistical significance for the difference in mean SBD coverage estimates was produced by paired t-test [17] for Benin and Malawi where the cluster means were normally distributed, and by Wilcoxon-Signed-Rank [18] test in India where the cluster means were distributed significantly different from normal according to a Shapiro-Wilk normality test [19].

Catchment area analysis

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

Table 1. Proportion of School-Age Children in DeWorm3 Catchment Areas with Unmatched School 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

https://doi.org/10.1371/journal.pntd.0010401.t001

mapped in QGIS, along with the residential locations of each child for whom self-reported school attendance was available.

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

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 minimum-bounding geographies that included the area within which SAC reported attending each school. A naïve version of this analysis was conducted using all SAC whose responses were matched to a specific school.

In order to obtain more discriminatory catchment areas for use in further analysis, the catchment area creation process was repeated after removing statistical outliers by furthest geographical distance between SAC and their self-reported schools. Using the QGIS processing toolbox function "Join by Lines (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).

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.

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 intervention arms was calculated by the DeWorm3 trial team during the course of previously published analyses [9]. These calculations resulted in a minimum site sample size of 80,000 individuals, with individual clusters containing a minimum of 1,650 individuals.

Results

Cluster-level SAC SBD coverages were consistently lower than the WHO 75% coverage goal for SAC and PSAC [20], with lower than 50% SBD coverage in 3 of 20 intervention clusters in India, 19 of 20 in Benin, and all 20 intervention clusters in Malawi (Fig 2).

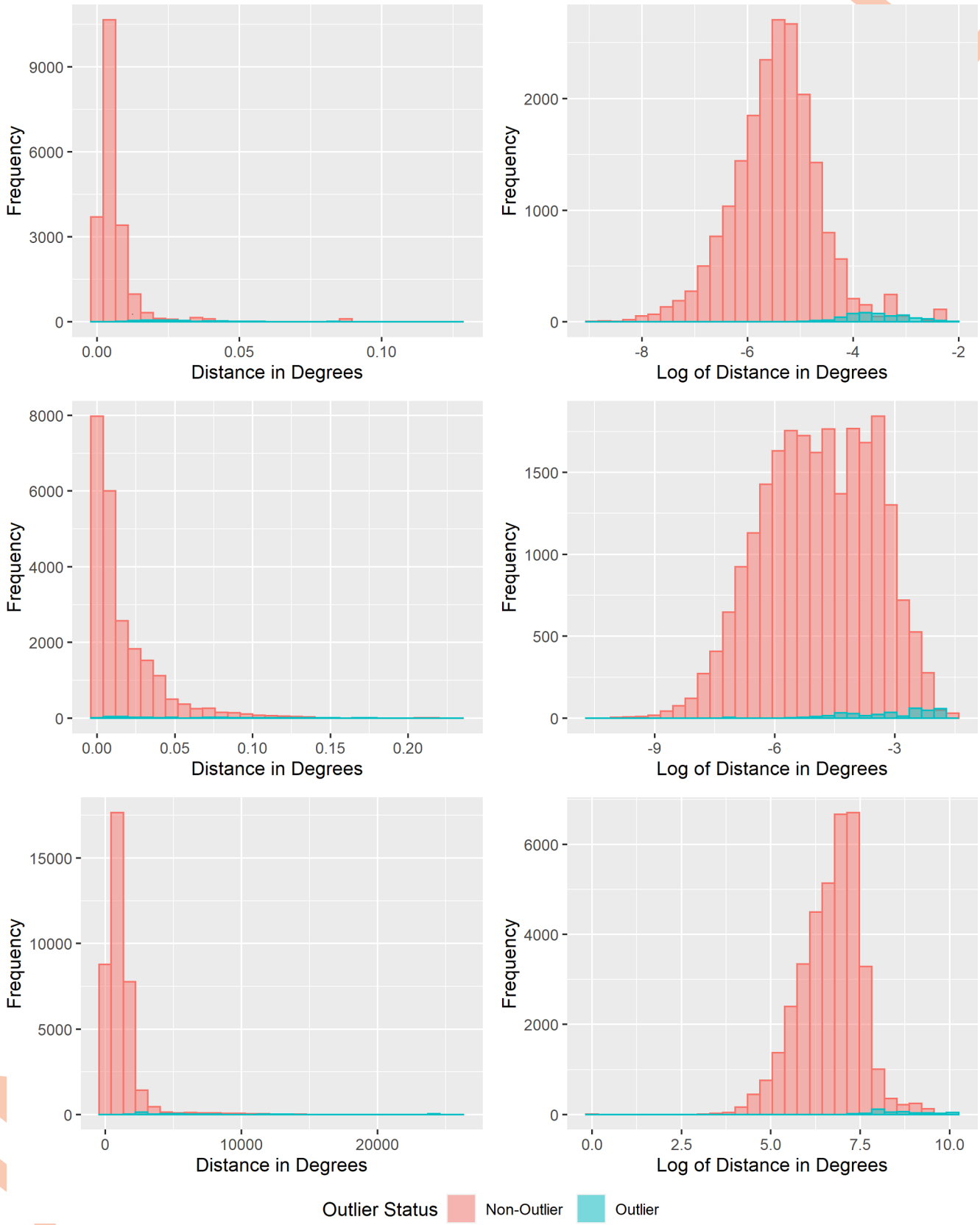
Frequency
Distance in Degrees
Log of Distance in Degrees
Outlier Status
Non-Outlier
Outlier

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.

https://doi.org/10.1371/journal.pntd.0010401.g001

These errors were 0.096 in Malawi, 0.135 in India, and 0.209 in Benin (Fig 3).

Applying these MSE measurements to the control arm school-level SBD coverages resulted in projected individual-level SBD estimates where none were previously available (Fig 4).

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

The catchment areas created in this analysis show a high degree of overlap, particularly in the highly 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 SAC. Recalculated school catchment areas were shown to overlap each other less than in the naïve analysis, particularly in Benin (Fig 6A).

This reflects the reality that school catchment areas are geographically indistinct and do not merely include the SAC closest to them. This means that any attempt to fix catchment areas solely using geographic proximity is unlikely to be accurate, especially in densely populated urban centers with many schools. For example, of the 40 clusters comprising the Benin study area, in only 7 clusters did more than 50% of SAC attend the school that was closest to their home by Euclidean distance, and in only one cluster was this the case for more than 65% of SAC (S1 Table). Overall, the percentage of 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).

Discussion

The success of global STH programs is critically dependent on the accurate measurement of programmatic treatment coverage. Funding mechanisms, policy directives, and the support of local stakeholders all require that programs can demonstrate the reach and the impact that they provide to local communities. Accurate reporting of data is of fundamental importance to achieving this goal. This paper seeks to enhance our understanding of deworming coverage provided during SBD by demonstrating the error inherent in school-level data collection when compared to a gold standard of individual-level reporting. This paper compares individual-level SBD data collected via community-wide MDA from the DeWorm3 trial's intervention arm against school-level SBD data collected via school-level reporting in both arms to calculate the difference between these estimates in the intervention arm clusters where both numbers are available. This is a new approach for characterizing discrepancies of 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.

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 over-estimating SAC deworming coverage when compared to a gold standard of individual-level 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 trial such as DeWorm3. If school-level data

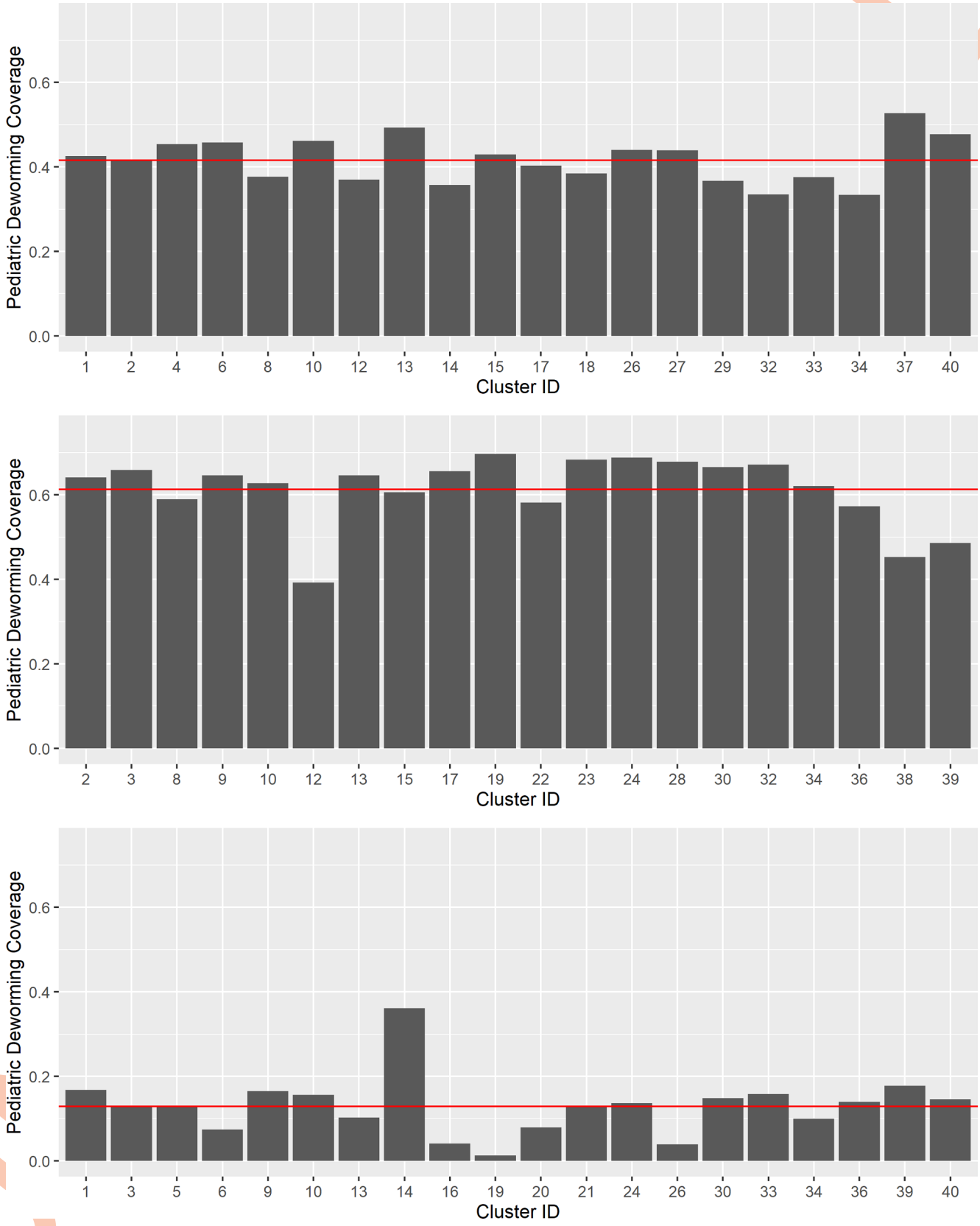

Fig 2. Individual-Level Reporting of SBD in Intervention Clusters, by Study Site. (A) Benin. (B) India. (C) Malawi. Mean SAC Deworming Coverage shown as a red slicer line. Y-Axis set to WHO target of 75% coverage for SAC and PSAC.

https://doi.org/10.1371/journal.pntd.0010401.g002

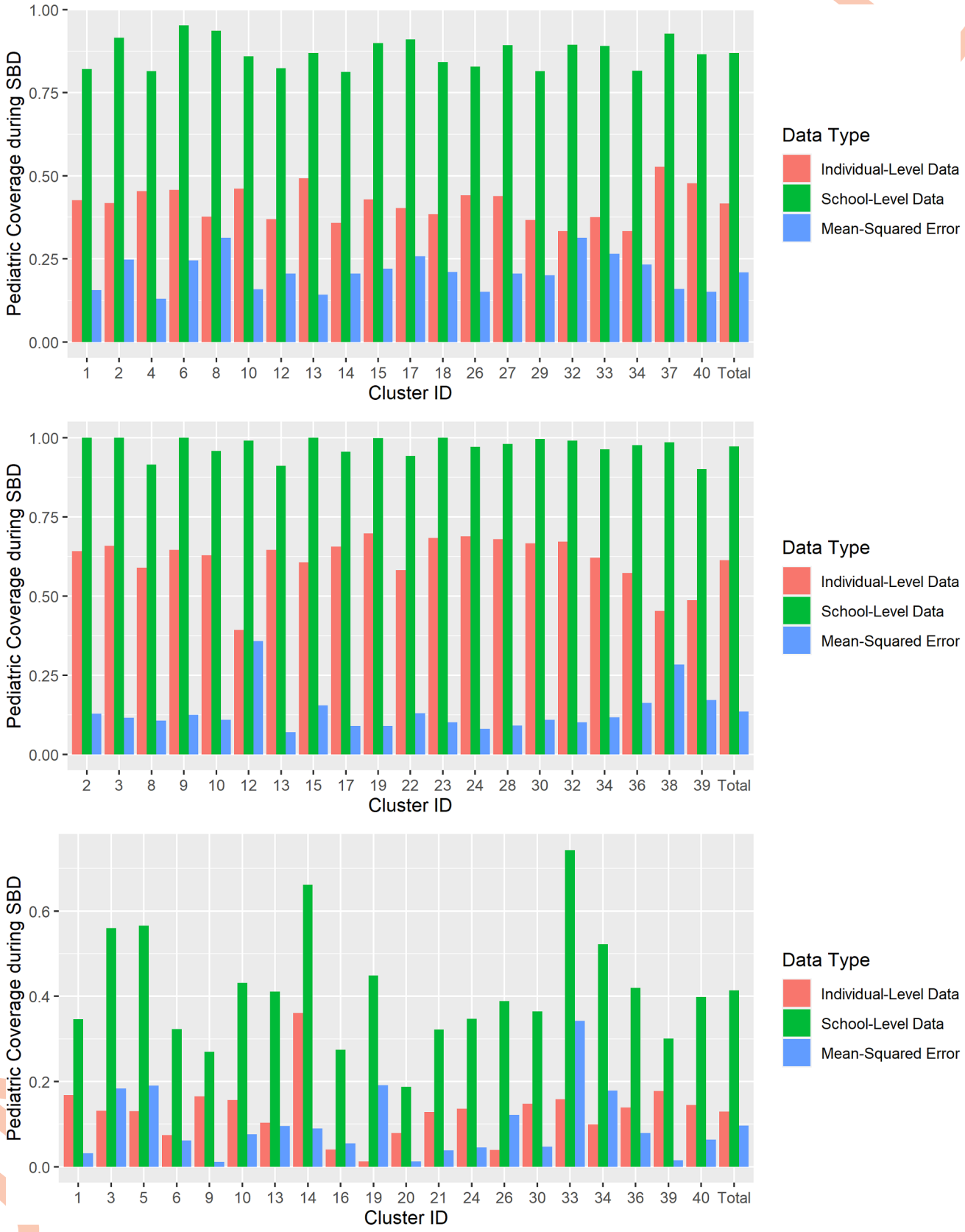
Pediatric Coverage during SBD
1.00
0.75
0.50
0.25
0.00
1 2 4 6 8 10 12 13 14 15 17 18 26 27 29 32 33 34 37 40 Total
Cluster ID
Data Type
Individual-Level Data
School-Level Data
Mean-Squared Error
Pediatric Coverage during SBD
1.00
0.75
0.50
0.25
0.00
2 3 8 9 10 12 13 15 17 19 22 23 24 28 30 32 34 36 38 39 Total
Cluster ID
Data Type
Individual-Level Data
School-Level Data
Mean-Squared Error
Pediatric Coverage during SBD
0.6
0.4
0.2
0.0
1 3 5 6 9 10 13 14 16 19 20 21 24 26 30 33 34 36 39 40 Total
Cluster ID
Data Type
Individual-Level Data
School-Level Data
Mean-Squared Error

Fig 3. Comparison of cluster-level SAC SBD coverage from individual-level data collection to school-level SBD in intervention clusters, by site. (A) Benin. (B) India. (C) Malawi.

https://doi.org/10.1371/journal.pntd.0010401.g003

are consistently overestimating coverage, then decisions made using these data at the program level may be made under the assumption that SBD coverage is higher than it actually is. Such a discrepancy could result in miscalculation of the duration and frequency of deworming for future planning, premature cessation of MDA or disruptive allocation of resources based on a flawed assumption of treatment coverage.

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

The finding that coverage reported by schools or district health authorities is consistently higher than coverage as determined by individual-level survey responses has been previously observed in trials [21–23], and when comparing national reporting of school-based treatment to WHO versus individual-level Demographic and Health Survey results [24]. Previous papers have postulated that the use of school or district-level reported coverages may open studies to biased results that overestimate deworming coverage due to exaggeration by community drug distributors, local health officials, or schoolteachers [11]. Another standard concern for survey-based data collection is recall bias, although this is unlikely to have affected the low coverages reported during individual-level data collection, as the DeWorm3 trial utilized indelible ink to mark the fingers of SAC who were treated at school in order to mitigate this concern [25]. In discussions with DeWorm3 program staff, it was noted that SBD coverages derived from individual-level data collection for the data analyzed in this paper might have been unusually low in Malawi because of a longer-than-usual gap between SBD and community-wide drug 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 from the fingers of SAC, preventing the use of this marking for accurate measurement of previous SBD treatment.

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 [26], Bayesian modeling [27], and cost-distance approaches [28] to estimate likely catchment areas for service delivery

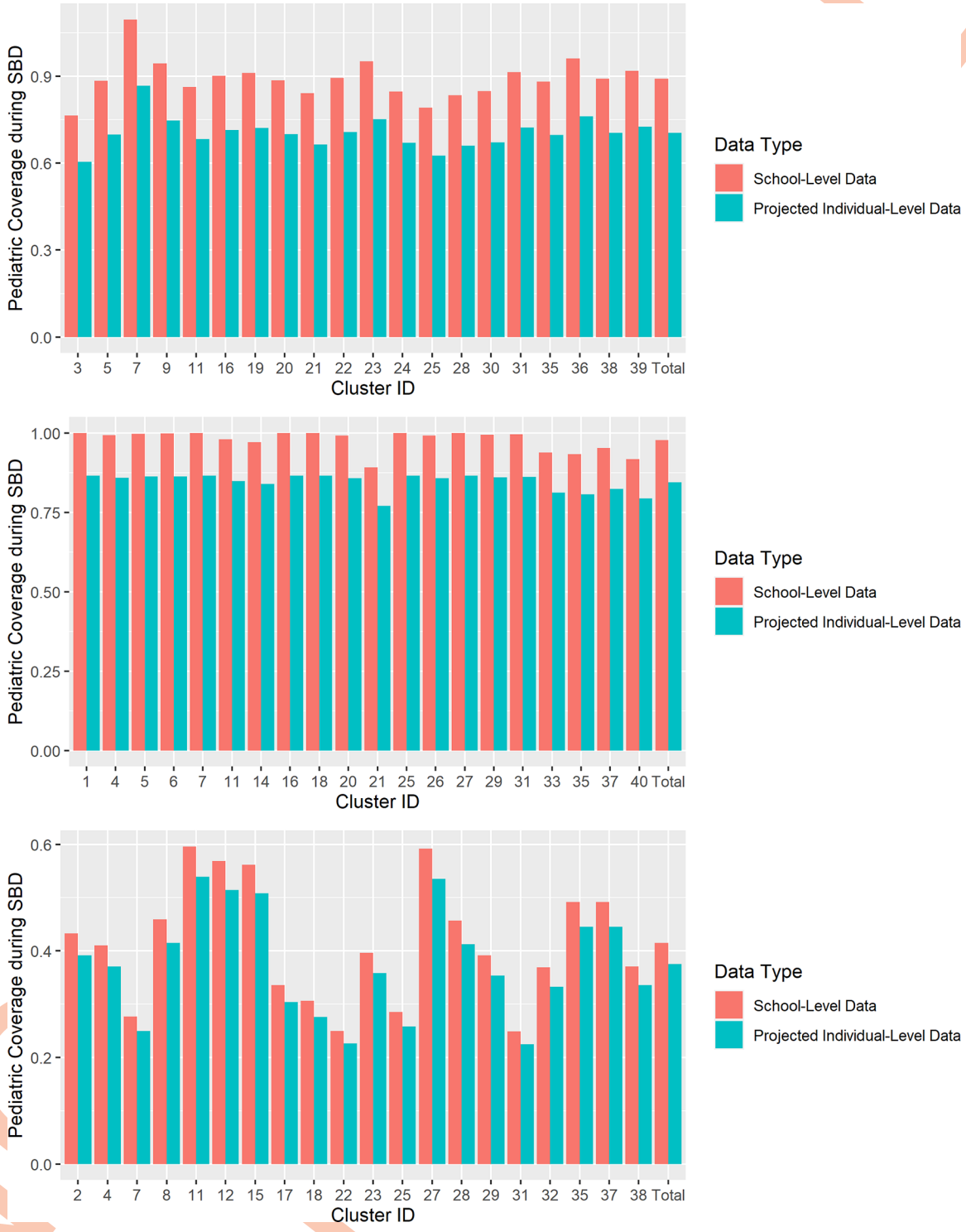

Fig 4. Comparison of school-level SBD coverage estimates to projected SBD coverage estimates in control clusters if individual-level data had been available, by study site. (A) Benin. (B) India. (C) Malawi.

https://doi.org/10.1371/journal.pntd.0010401.g004

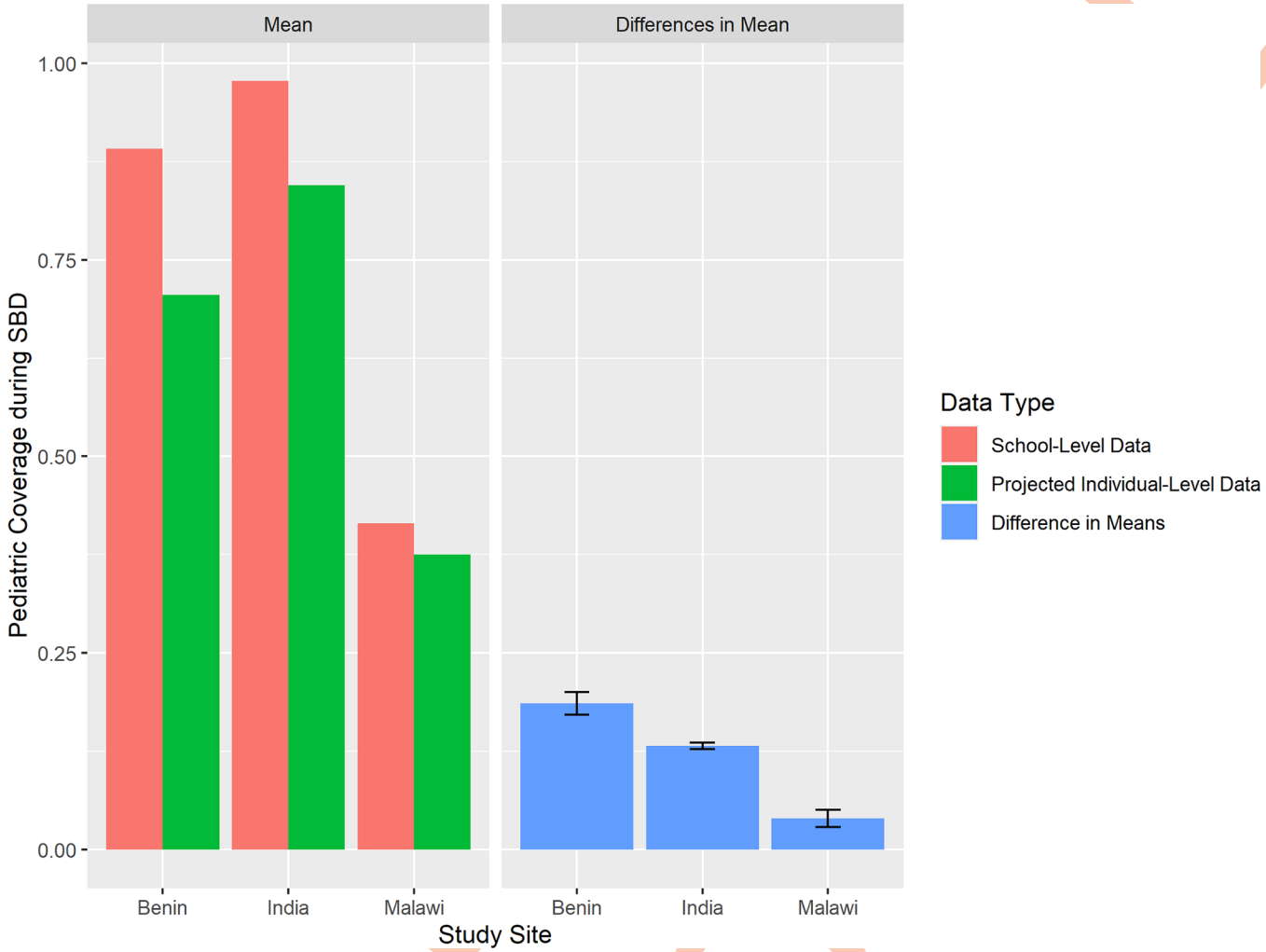

Fig 5. Comparison of school-level average SBD coverage to average individual-level SBD coverage in control clusters after application of mean-squared error rates.

https://doi.org/10.1371/journal.pntd.0010401.g005

of health posts, hospitals, and schools for deworming activities. The DeWorm3 trial provides access to geolocation information for each SAC treated within the study areas, allowing for exact visualization of reported catchment areas. The school catchment areas in the DeWorm3 trial areas overlap each other significantly, especially in the population centers where schools are geographically near to one another. In addition, most SAC in these study areas did not receive their deworming treatments at the school geographically nearest to them. If this is the case for populations in other study areas, this means that proximity-based approaches for defining health activity catchment areas may introduce considerable inaccuracies for determining relevant population denominators. A lack of clear communication to survey administrators regarding the geographic limits of clusters, municipal boundaries, and catchment areas has been previously identified as a possible driver of variation in deworming treatment coverage [11,29]. Therefore, a potential benefit stemming from the visualization of these catchment areas may include more streamlined data collection and drug-distribution practices for school-based and community-wide MDA in the DeWorm3 trial area.

The DeWorm3 trial is an exceptionally large and rigorously conducted study, with harmonized data collection practices across three different countries. Additionally, these data had

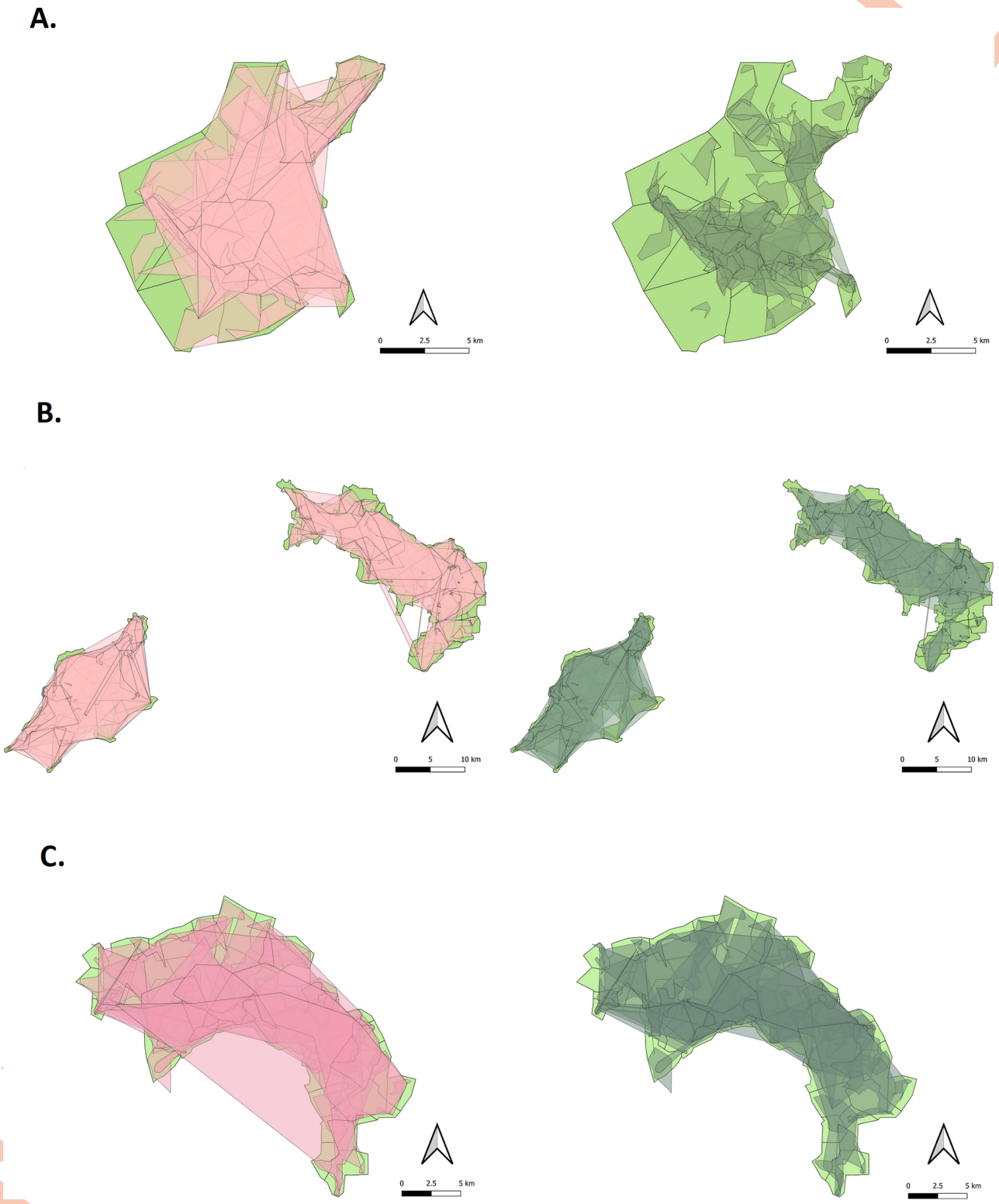

Fig 6. Catchment areas for each school by study area, created via QGIS Concave Hull (K-nearest neighbor) analysis. (A) Benin. (B) India. (C) Malawi. The left side of each panel shows naïve catchment areas in semi-transparent pink created prior to removal of geographic outliers, while the right side shows catchment areas in semi-transparent green created after removal of geographic outliers. Both overlay original cluster boundaries in green. Figures created using QGIS 3.10.7 A Coruña.

https://doi.org/10.1371/journal.pntd.0010401.g006

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

https://doi.org/10.1371/journal.pntd.0010401.t002

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

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 would likely go for deworming treatments in a given cluster. And for the assignment of population-based weights, the self-reported attendance data do not allow for consideration of those SAC who may not have been reached during data collection, or whose responses were matched to the incorrect school.

Free-text school name entries also pose a limitation because there is no perfect recording of where each child in a given cluster received their SBD treatment. To minimize the effect of this limitation, thousands of free-text entries were analyzed and matched, where possible, to the appropriate school in each study site's school registry, via the iterative process described in the methods section. In this analysis, 17% of SAC respondents in Benin were not matched to a specific school, while 15% and 10% of SAC remained unmatched to a school in India and Malawi, respectively. These unmatched SAC reported going to schools that were unknown, were outside the study area, or were schools for which coverage levels and enrollment data were not available, with informal or nursery schools accounting for the largest percentage of free-text entries that cannot be matched to a registered school. As an illustrative example, it is believed that over 97% of the 3,929 unmatched free text entries for the India study site of the DeWorm3 trial refer to Anganwadi (informal government-run education centers for rural childcare developed under the Integrated Child Development Scheme) centers due to similarities in the submitted free-text responses. This has been a limitation in previous studies, including in Bangladesh, where informal schooling complicated calculation of coverage and population statistics [30]. If large numbers of these unmatched respondents actually attended one of the schools in the school registry, then the weighted average formula for determining coverage in SBD from school-level data would have slightly underweighted the schools that they attended. This would only provide a significant limitation if large numbers of SAC were assumed to attend an erroneous school in the registry, leading to a slight overweighting of that school in the weighted average coverage formula.

Finally, it is important to note that this analysis only utilized data collected in year two of the DeWorm3 trial, which consisted of two treatment rounds in India and 1–2 rounds in Benin and Malawi, as these data were determined by the DeWorm3 program staff to be the most complete dataset available. In year one, data were less reliable because certain data collection processes had not yet been standardized, and in year three, data collection was disrupted by the worldwide emergence of SARS-CoV-2. A more complete analysis would examine data across multiple study years to determine if SAC coverage rates fluctuated significantly from year to year, and if observed error in the school-level SBD data were sustained throughout the duration of the trial.

As programs transition to targeting elimination of STH as a public health problem, accurate coverage determinations become increasingly important. Studies employing methods to assess coverage based on school-based reporting may need to adjust their reporting methods to account for the possibility of overreporting. Where it is not possible to do supplemental individual-level community coverage surveys due to budgetary or logistical constraints, acknowledgement should be made of possible overestimation by school-based reporting.

Conclusion

This analysis utilized data from the DeWorm3 trial to quantify discrepancies between cluster-level estimates of SAC deworming coverage derived from gold-standard individual-level data collection during community-wide MDA in the intervention arm and historically less reliable school-level data collection during the delivery of standard of care school-based deworming in both trial arms. These estimates, derived from three different country sites and across many settings within those countries, indicate that school-level reporting of SAC deworming coverage likely overestimate program coverage in these age groups. These results suggest that school-level reporting of SAC deworming coverage should be validated by individual-level data collection. Novel interventions to improve data collection within MDA programs are needed to ensure that accurate reporting informs programmatic decision-making and resource allocation.

Supporting information

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

Acknowledgments

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.

Author Contributions

Conceptualization: William Sheahan, Arianna Rubin Means, Judd L. Walson, Kristjana Hrönn Ásbjörnsdóttir.

Data curation: William Sheahan, Sean Galagan, Gideon John Israel, Venkateshprabhu Janagaraj, Chloe Morozoff, Emily Pearman, Rohan Michael Ramesh, Amy Roll, Alexandra Schaefer, James Simwanza, Stefan Witek-McManus.

Formal analysis: William Sheahan, Saravanakumar Puthupalayam Kaliappan, Kristjana Hrönn Ásbjörnsdóttir.

Funding acquisition: Judd L. Walson.

Investigation: Roy Anderson, Kumudha Aruldas, Euripide Avokpaho, Parfait Houngbegnon, Gideon John Israel, Venkateshprabhu Janagaraj, Saravanakumar Puthupalayam Kaliappan, Rohan Michael Ramesh, James Simwanza, Stefan Witek-McManus, Robin Bailey, Moudachirou Ibikounlé, Khumbo Kalua, Adrian J. F. Luty, Rachel Pullan, Kristjana Hrönn Ásbjörnsdóttir.

Methodology: William Sheahan, Roy Anderson, Kumudha Aruldas, Saravanakumar Puthupalayam Kaliappan, Sitara S. R. Ajjampur, Adrian J. F. Luty, Rachel Pullan, Kristjana Hrönn Ásbjörnsdóttir.

Project administration: Jeanne Goodman, Parfait Houngbegnon, Sitara S. R. Ajjampur, Khumbo Kalua, Adrian J. F. Luty, Judd L. Walson.

Resources: Rachel Pullan, Judd L. Walson.

Software: William Sheahan.

Supervision: Kumudha Aruldas, Sean Galagan, Saravanakumar Puthupalayam Kaliappan, Arianna Rubin Means, Sitara S. R. Ajjampur, Moudachirou Ibikounlé, Judd L. Walson, Kristjana Hrönn Ásbjörnsdóttir.

Validation: Sean Galagan.

Visualization: William Sheahan.

Writing – original draft: William Sheahan.

Writing – review & editing: William Sheahan, Kumudha Aruldas, Euripide Avokpaho, Parfait Houngbegnon, Gideon John Israel, Venkateshprabhu Janagaraj, Saravanakumar Puthupalayam Kaliappan, Rohan Michael Ramesh, Sitara S. R. Ajjampur, Robin Bailey, Moudachirou Ibikounlé, Adrian J. F. Luty, Judd L. Walson, Kristjana Hrönn Ásbjörnsdóttir.

References

1. Soil-transmitted helminth infections [Internet]. World Health Organization. 2020 [cited 2021 Jun 10]. Available from: https://www.who.int/news-room/fact-sheets/detail/soil-transmitted-helminth-infections
2. 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. https://doi.org/10.1016/S0140-6736(20)30925-9 PMID: 33069326
3. Coverage Evaluation Guidelines Final Draft_Nov 2016.pdf [Internet]. [cited 2021 Jun 10]. Available from: https://www.ntdsupport.org/sites/default/files/uploads/docs/resources/Coverage%20Evaluation%20Guidelines%20Final%20Draft_Nov%202016.pdf
4. Manyeh AK, Ibisomi L, Ramaswamy R, Baiden F, Chirwa T. Exploring factors affecting quality implementation of lymphatic filariasis mass drug administration in Bole and Central Gonja Districts in Northern Ghana. PLoS Negl Trop Dis. 2020 Aug 17; 14(8):e0007009. https://doi.org/10.1371/journal.pntd.0007009 PMID: 32804967
5. 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. https://doi.org/10.1371/journal.pntd.0006043 PMID: 29077723
6. Oshish A, AlKohlani A, Hamed A, Kamel N, AlSoofi A, Farouk H, et al. Towards nationwide control of schistosomiasis in Yemen: a pilot project to expand treatment to the whole community. Trans R Soc Trop Med Hyg. 2011; 105(11):617–627. Epub 2011/09/13. https://doi.org/10.1016/j.trstmh.2011.07.013 PMID: 21907376
7. Gabrielli AF, Toure S, Sellin B, Sellin E, Ky C, Ouedraogo H, et al. A combined school- and community-based campaign targeting all school-age children of Burkina Faso against schistosomiasis and soil-transmitted helminthiasis: performance, financial costs and implications for sustainability. Acta Trop. 2006; 99(2–3):234–242. https://doi.org/10.1016/j.actatropica.2006.08.008 PMID: 16997268

8. Dembélé M, Bamani S, Dembélé R, Traoré MO, Goita S, Traoré MN, et al. Implementing preventive chemotherapy through an integrated national neglected tropical disease control program in Mali. PLoS Negl Trop Dis. 2012; 6(3):e1574. https://doi.org/10.1371/journal.pntd.0001574 PMID: 22448294
9. 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. https://doi.org/10.1371/journal.pntd.0006166 PMID: 29346377
10. The DeWorm3 Trials Team (2020) 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 Negl Trop Dis 14(11): e0008771. https://doi.org/10.1371/journal.pntd.0008771 PMID: 33137100
11. 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. https://doi.org/10.1016/j.ijpara.2016.01.006 PMID: 26940547
12. 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. https://doi.org/10.1371/journal.pntd.0008771 PMID: 33137100
13. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.
14. QGIS Development Team, 2021. QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://qgis.org
15. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
16. Zheng—Methods of Evaluating Estimators.pdf [Internet]. [cited 2021 Jun 10]. Available from: http://people.missouristate.edu/songfengzheng/teaching/mth541/lecture%20notes/evaluation.pdf
17. Ross A., Willson V.L. (2017) Paired Samples T-Test. In: Basic and Advanced Statistical Tests. SensePublishers, Rotterdam.
18. Rey D., Neuhäuser M. (2011) Wilcoxon-Signed-Rank Test. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg.
19. Shapiro S.S. & Wilk M.B. (1965) An analysis of variance for normality (complete samples). Biometrika, Vol. 52, No. 3/4.
20. World Health Organization, Nutrition for Health and Development, World Health Organization, Department of Control of Neglected Tropical Diseases, World Health Organization. Preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups: guideline. [Internet]. 2017 [cited 2021 Jun 10]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK487927/
21. Katabarwa MN, Griswold E, Habomugisha P, Eyamba A, Byamukama E, Nwane P, et al. Comparison of Reported and Survey-Based Coverage in Onchocerciasis Programs over a Period of 8 Years in Cameroon and Uganda. The American Journal of Tropical Medicine and Hygiene. 2019 Mar 25; 100 (5):1208–15. https://doi.org/10.4269/ajtmh.18-0680 PMID: 30915956
22. Worrell C, Mathieu E. Drug coverage surveys for neglected tropical diseases: 10 years of field experience. Am J Trop Med Hyg. 2012 Aug; 87(2):216–22. https://doi.org/10.4269/ajtmh.2012.12-0167 PMID: 22855750
23. Binder S, Campbell CH, Castleman JD, Kittur N, Kinung'hi SM, Olsen A, et al. Lessons Learned in Conducting Mass Drug Administration for Schistosomiasis Control and Measuring Coverage in an Operational Research Setting. Am J Trop Med Hyg. 2020 Jul; 103(1 Suppl):105–13. https://doi.org/10.4269/ajtmh.19-0789 PMID: 32400352
24. Lo NC, Gupta R, Addiss DG, Bendavid E, Heft-Neal S, Mikhailov A, et al. PLOS Comparison of World Health Organization and Demographic and Health Surveys data to estimate sub-national deworming coverage in pre-school aged children. Neglected Tropical Diseases. 2020 Aug 17; 14(8):e0008551.
25. Cromwell EA, Ngondi J, McFarland D, King JD, Emerson PM. Methods for estimating population coverage of mass distribution programmes: a review of practices in relation to trachoma control. Transactions of The Royal Society of Tropical Medicine and Hygiene. 2012 Oct 1; 106(10):588–95. https://doi.org/10.1016/j.trstmh.2012.07.011 PMID: 22884927
26. Cattaneo A, Nelson A, McMenomy T. Global mapping of urban–rural catchment areas reveals unequal access to services. PNAS [Internet]. 2021 Jan 12 [cited 2021 Jun 10];118(2). Available from: http://www.pnas.org/content/118/2/e2011990118 https://doi.org/10.1073/pnas.2011990118 PMID: 33431572

27. Alegana VA, Khazenzi C, Akech SO, Snow RW. Estimating hospital catchments from in-patient admission records: a spatial statistical approach applied to malaria. Sci Rep. 2020 Jan 28; 10(1):1324. https://doi.org/10.1038/s41598-020-58284-0 PMID: 31992809

28. Stresman GH, Stevenson JC, Owaga C, Marube E, Anyango C, Drakeley C, et al. Validation of three geolocation strategies for health-facility attendees for research and public health surveillance in a rural setting in western Kenya. Epidemiology & Infection. 2014 Sep; 142(9):1978–89.

29. Pullan RL, Halliday KE, Oswald WE, Mcharo C, Beaumont E, Kepha S, et al. Effects, equity, and cost of school-based and community-wide treatment strategies for soil-transmitted helminths in Kenya: a cluster-randomised controlled trial. The Lancet. 2019 May 18; 393(10185):2039–50.

30. Nath TC, Padmawati RS, Murhandarwati EH. Barriers and gaps in utilization and coverage of mass drug administration program against soil-transmitted helminth infection in Bangladesh: An implementation research. Journal of Infection and Public Health. 2019 Mar 1; 12(2):205–12. https://doi.org/10.1016/j.jiph.2018.10.002 PMID: 30385237