

Investigating Discrepancies in School-Based Deworming Coverage Estimates from the DeWorm3 Trial

William Sheahan

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2021

Committee:

Judd Walson

Kristjana Asbjornsdottir

Program Authorized to Offer Degree:

Epidemiology

Copyright © 2021

William Sheahan

University of Washington

**Abstract**

Investigating Discrepancies in School-Based Deworming Coverage Estimates from the DeWorm3 Trial

William Sheahan

Chair of the Supervisory Committee:

Judd Walson

Global Health, Adjunct Epidemiology

Background: Soil Transmitted Helminths (STH) infect over 1.5 billion people globally, causing anemia and growth stunting that result in an annual toll of 1.9 million Disability-Adjusted Life Years. School-based deworming (SBD) via Mass Drug Administration (MDA) campaigns with albendazole or mebendazole has been recommended by the World Health Organization to reduce the prevalence of STH in endemic areas. The DeWorm3 Trial is a cluster randomized trial investigating the feasibility of interrupting the transmission of STH with community-wide MDA for three study sites in Benin, India, and Malawi.

Methodology/Principal Findings: 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 available from the DeWorm3 Trial's intervention arm. Population-weighted averages of school-level SBD were calculated at the cluster level and compared to directly aggregated individual-level estimates of SBD to produce a Mean Squared Error (MSE) estimate for each study site. These MSE values were applied to SBD estimates from the control arm of the DeWorm3 Trial, where only school-

level reporting of SBD had been collected, in order to project what coverages would have been for individual-level reporting of SBD.

In each site, SBD coverage was reported as being substantially higher in the school-level datasets than in the gold standard individual-level datasets, indicating that school-level SBD reporting may be over-estimating SBD coverage. In Benin, average SBD coverage in the control arm dropped from 89.1% to 70.5% when applying observed MSE to project expected coverages. In India, the expected coverages decreased from 97.7% to 84.5%, and in Malawi, expected coverages decreased from 41.5% to 37.5%.

Conclusions/Significance: These estimates indicate that school-level reporting of SBD likely overestimate program reach and effectiveness. These findings suggest that current estimates of coverage derived from school-based program data may substantially overestimate true pediatric deworming coverage.

## 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 growth stunting resulting in a toll of almost 2 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 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. 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 and may be at even higher risk for STH infection. To address this gap, and to 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 in concert with school-based MDA (4).

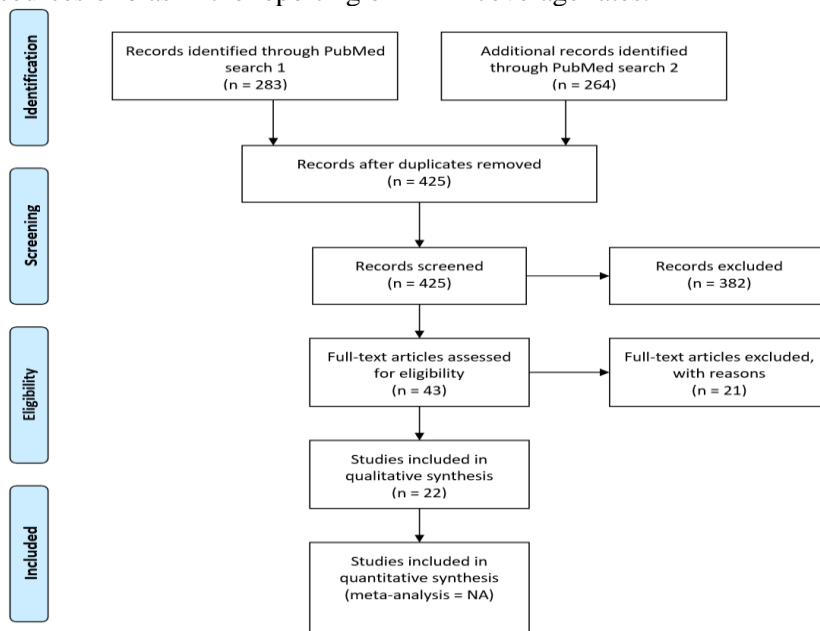
The DeWorm3 Trial is a cluster randomized trial examining the feasibility of interrupting the transmission of STH through MDA, comparing community-wide MDA with standard of care school-based MDA in endemic areas of Benin, India, and Malawi. This analysis uses DeWorm3 Trial data to examine the degree to which bias in school-level reporting of MDA results may lead to overestimation of SBD coverage when compared with coverage estimates from individual-level data collection. 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 the number of children who attend school on deworming days, rather than the overall population of children living in targeted communities. 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 (5).

DeWorm3 provides a novel opportunity to analyze these SAC coverage data due to the rigorous and systematic collection 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 the bias inherent in traditional school-based coverage estimates. This comparison of school-level data to individual-level reporting has broad implications for understanding bias in reported program coverage for school-based mass drug administration programs.

## Methods

Literature Review: A preliminary literature review was conducted to inform the researcher's background in these contexts, with the primary aim of identifying factors that have previously been reported to account for bias in the estimation of coverage levels for Mass Drug Administration. In particular, the review sought to isolate factors that have been relevant in reporting coverage estimates for deworming activities, including campaigns to treat soil-transmitted helminths, lymphatic filariasis, schistosomiasis, onchocerciasis, trachoma, and other Neglected Tropical Diseases.

Figure 3: A PRISMA diagram depicting search results and screening for the preliminary literature review on sources of bias in the reporting of MDA coverage rates.



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. [PLoS Med 6\(7\): e1000097](https://doi.org/10.1371/journal.pmed1000097). doi:10.1371/journal.pmed1000097

Search strings created for this review in PubMed and executed on February 8<sup>th</sup>, 2021 are included in the appendices of this paper. An initial search returned 425 deduplicated results that were examined for relevance in the title/abstract review stage. 43 studies were deemed relevant and advanced to a full-text

review stage, at which point 21 studies were eliminated for lack of relevance due to factors, including but not limited to: a failure to mention specific methods of coverage calculation beyond a quotient of the percent treated over the percent eligible, or a failure to list any suggestions for drivers of coverage over- or under-reporting. From this literature review, a number of possible drivers of bias in coverage estimation for mass drug administration were identified. These drivers of bias will be considered in the discussion section of this paper.

Study Design: All data used in this analysis come from the DeWorm3 trial. This is a cluster-randomized trial examining the feasibility of interrupting the transmission of soil-transmitted helminths through mass drug administration and a detailed description of the trial methods have been reported elsewhere (6). 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.

Study Setting: The DeWorm3 trial operates in three primary locations. In Benin, the Institut Recherche Clinique du Benin and Institut de Recherche pour le Developpement 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 primarily by the Christian Medical College, Vellore, in the state of Tamil Nadu (6). Each country included an entire administrative area (two administrative areas in India) inclusive of more than 80,000 individuals in each country. These administrative areas were divided into 40 total clusters, each containing a minimum of 1,650 individuals. Twenty of these clusters are randomized to receive community-wide MDA, while twenty receive standard of care SBD.

Ethics statement: The DeWorm3 Project 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) from 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 overall study was approved by The Human Subjects Division at the University of Washington (STUDY00000180). This particular analysis also 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. Written informed consent was sought if the individual could write, while for any individual who could not write, oral consent was given in the presence of a witness and documented with a thumbprint. The use of oral consent in these circumstances has been approved by the relevant ethical committees.

Study Subjects: As described in [Ásbjörnsdóttir et al, 2018](#), the DeWorm3 Trial tests the feasibility of interrupting STH transmission throughout selected communities, using 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 ([7](#)) that enumerates all households and individuals within each study area, and gives considerable detail to the demographic characteristics of each site.

Data Collection: No new data were collected during the course of this analysis. Previous collection of data through the DeWorm3 trial is detailed extensively in the original cluster randomized trial protocol ([6](#)). The DeWorm3 Trial has provided door-to-door community-wide MDA for soil-transmitted helminths in intervention arm clusters, timed to follow standard of care school-based MDA, while control

clusters continue to receive standard of care school-based MDA only. School-based MDA is implemented by the DeWorm3 team in Malawi, and without DeWorm3 involvement in India and Benin. 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, the control arm data are limited to routine school-level coverage data.

Data Analysis: Data were analyzed using two free and open-source software programs, R Studio (8) and QGIS (9).

Individual-level SBD coverage estimates were summarized to the cluster level based on the location of each respondent's home. Because SAC are 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 times 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 pediatric coverage estimates are available to quantify the discrepancy between these data sources. The individual-level data from community-wide drug distribution 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, which are presumed to be of higher quality, 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 bias from the mean square 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  $\theta$  (10), here taken to be the true pediatric coverage at the cluster-level which is approximated by individual-level coverage estimates from the intervention arm. If the calculated bias value is equal to zero, then we would assume the cluster-level estimates from the control arm of the Trial to be unbiased in comparison to the individual-level coverage estimates from the intervention arm.

After calculating the expected levels of bias in existing school-level coverage estimates from the intervention arm, the overall site bias rates were applied to control clusters where only school-level estimates were available. Multiplying the existing school-level estimates by the ascertained level of bias and subtracting this from the school-level estimate allowed us to project what the likely individual-level drug distribution coverage estimates would have been if those data had been collected in the control arm of the Trial.

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 three study areas 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 individual 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 the majority of SAC chose their designated school 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. A non-zero proportion of free-text entries for each site remained unmatched and were excluded from the catchment area creation process (Table 2).

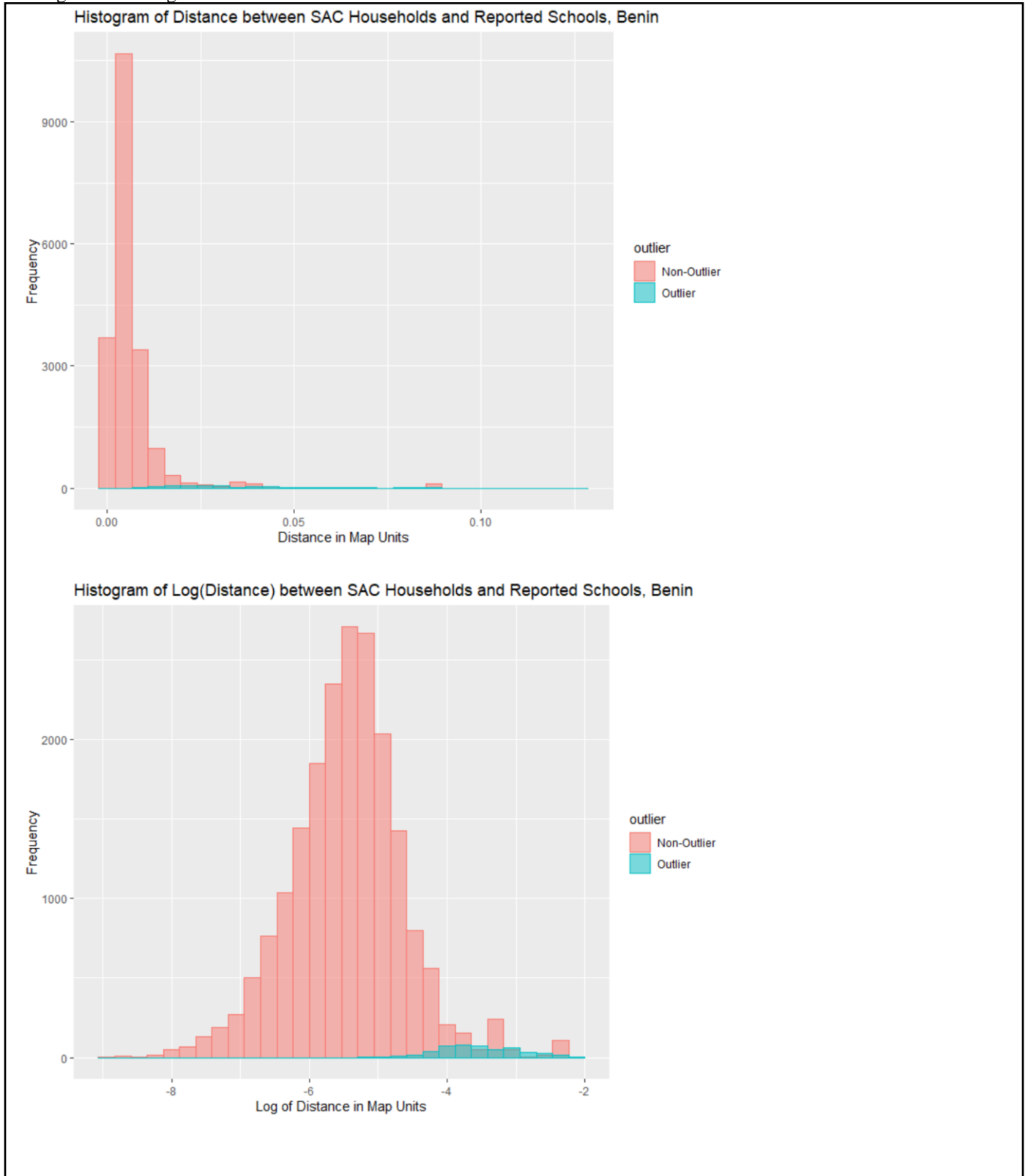
To obtain an initial understanding of the geographic range attributed to each school's catchment area, the QGIS processing toolbox function "Concave Hull (k-nearest neighbor)"<sup>1</sup> was utilized to create unique minimum-bounding geographies that included the area within which SAC reported attending 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 (Figure 1b).

Figure 1 a-b: Histograms of the distance between SAC and their self-reported schools. Figure 1a shows the distribution of the untransformed variable with distance measured in map units, while Figure 1b shows the histogram of the log-transformed variable.



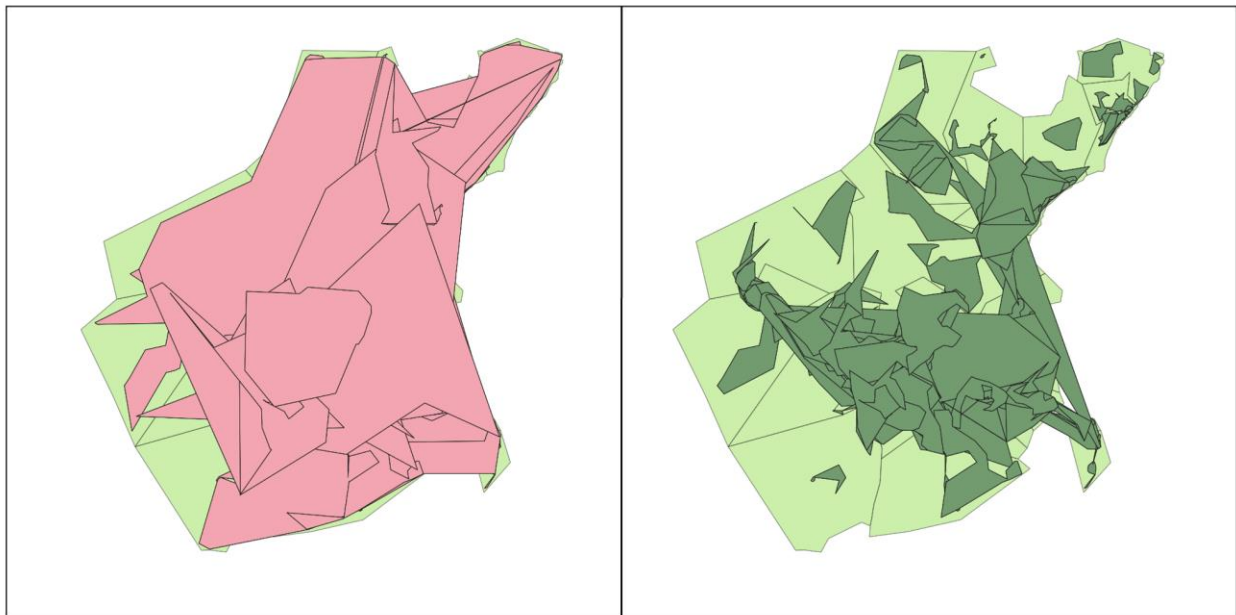
Because of the approximately log-normal distribution of the distance between the households and the reported schools of SAC in the DeWorm3 study areas, geographic outliers were removed for each school if the log-transformed distance between a respondent and their self-reported school was greater than 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.

Study Power: No new sample size calculations were conducted for this analysis, as the requisite statistical power has been calculated by the DeWorm3 trial team during the course of previously published analyses (6).

## Results

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 greatly distant SAC. Recalculated school catchment areas were shown to overlap each other less than in the naïve analysis, particularly in Benin ([Figure 2b](#)).

[Figure 2 a-b](#): Catchment areas for each school in the Benin study area, created via QGIS Concave Hull (K-nearest neighbor) analysis. Figure 2a shows naïve catchment areas created prior to removal of geographic outliers. Figure 2b shows catchment areas created after removal of geographic outliers.



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 geographically closest to their home, and in only one cluster was

this the case for more than 65% of SAC ([Supplemental Table 1](#)). Overall, fewer than one quarter of SAC attended the school geographically closest to their home in Benin and India, although in Malawi, that number was substantially higher at over 65% ([Table 1](#)).

**Table 1:**

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

Benin Proportion	Benin (n)	India Proportion	India (n)	Malawi Proportion	Malawi (n)
0.242	30199	0.127	28591	0.66	42583

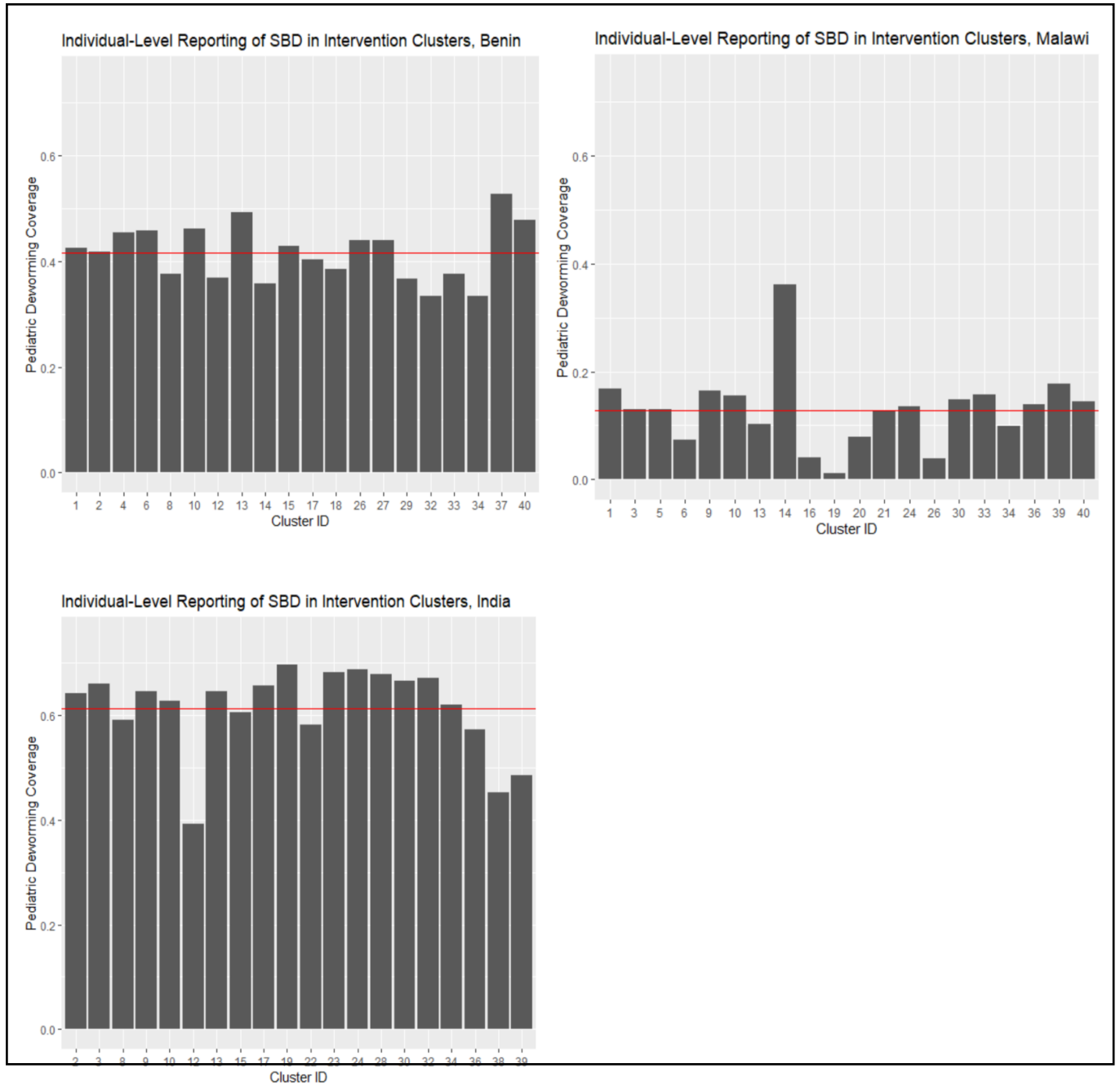
While acknowledging that the naïve catchment areas most accurately reflect self-reported school attendance for each study site, it is important to note that these include wide-ranging geographic outliers and therefore may not accurately predict where a hypothetical child for whom self-reported school attendance is unavailable would likely go for deworming treatments in a given cluster. A nonzero proportion of SAC in each study area were unable to be matched to a school in the DeWorm3 school registry ([Table 2](#)), either because they reported going to schools that were unknown, were outside the study area, or were informal nursery schools for which coverage levels and enrollment data were not available. In addition, 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.

Table 2:

Proportion of School-Age Children in DeWorm3 Catchment Areas with Unmatched Schoolnames, by Study Site

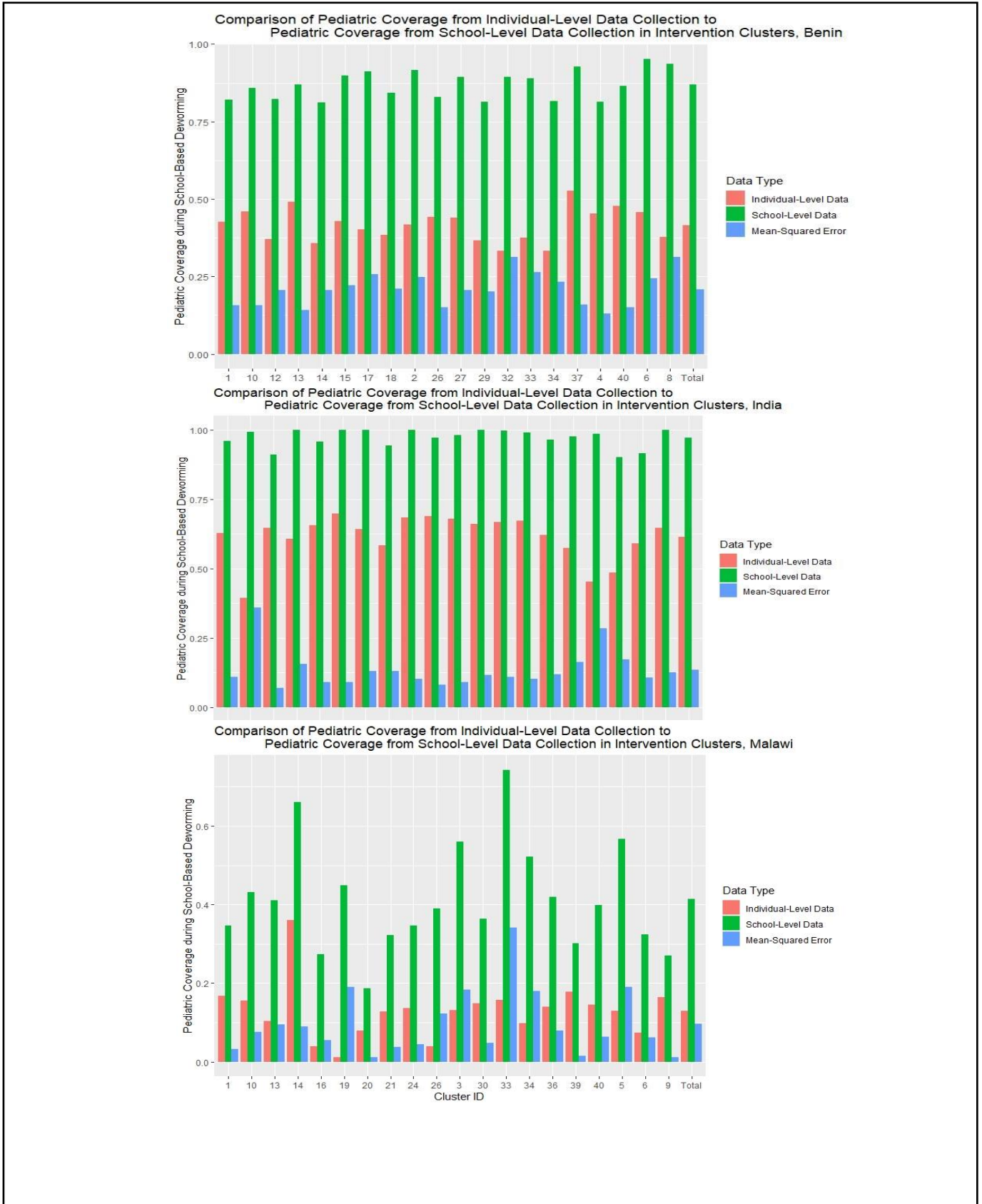
	Matched (n)	Unmatched (n)	Total (n)	Proportion Unmatched
Benin	20241	9958	30199	0.33
India	24262	4329	28591	0.15
Malawi	38339	4248	42587	0.10

**Figure 4 a-c:** Individual-Level Reporting of School-Based Deworming in Intervention Clusters, by Study Site. Mean Pediatric Deworming Coverage shown as red slicer line. Y-Axis set to WHO target of 75% coverage for SAC and PSAC.



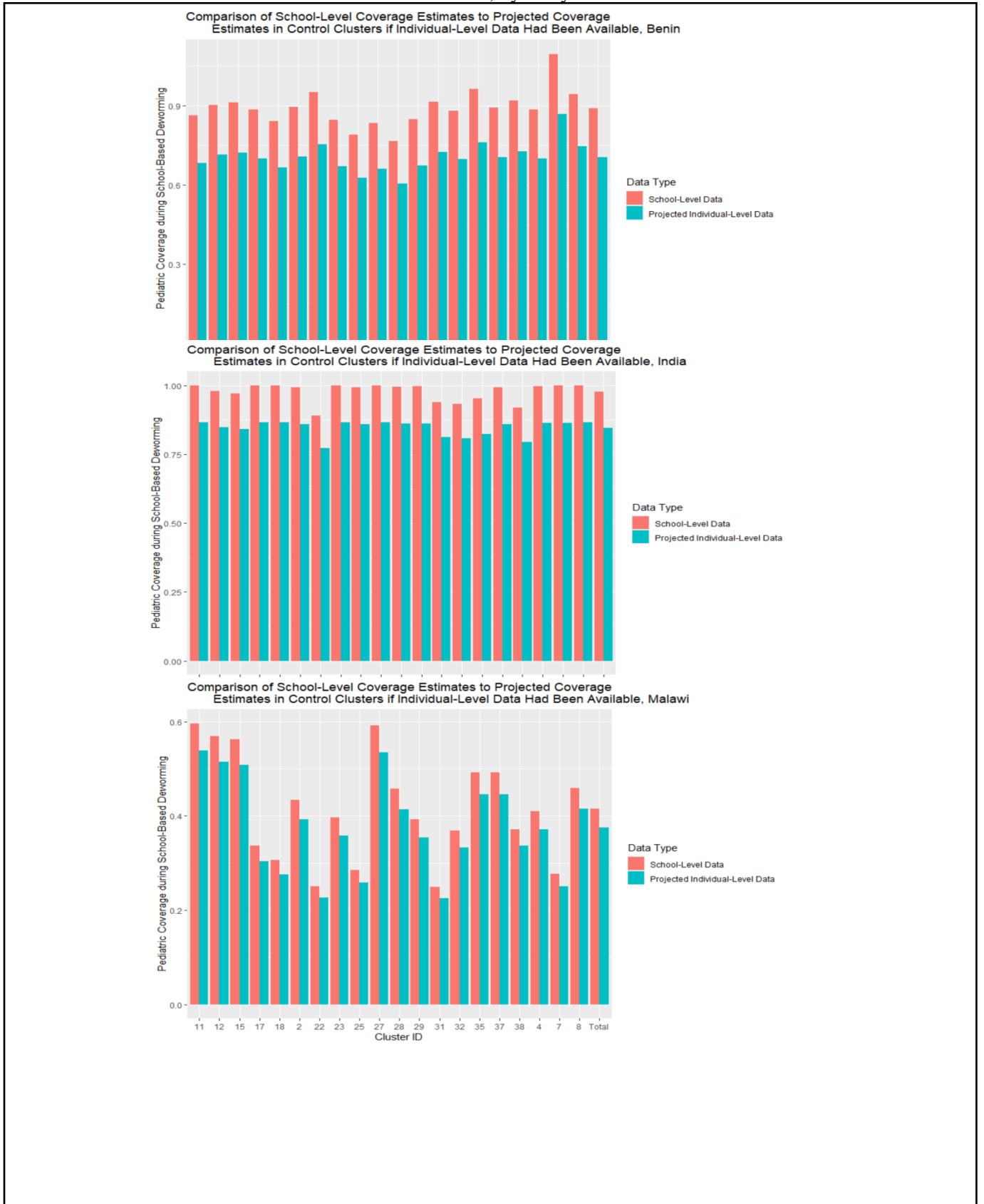
Cluster-level pediatric SBD coverages were calculated from the clusters randomized to receive the community-wide MDA intervention annually in Malawi and Benin, and bi-annually in India. These estimates specifically determined the cluster-level coverage as the proportion of SAC reached during community-wide MDA who reported being previously treated during SBD. These coverages were consistently lower than the WHO 75% coverage goal for SAC and PSAC ([11](#)), showing lower than 50% coverage in 3 of 20 intervention clusters in India, 19 of 20 in Benin, and all 20 intervention clusters in Malawi ([Figure 4](#)).

**Figure 5:** Comparison of cluster-level pediatric coverage from individual-level data collection to pediatric coverage from school-level data collection in intervention clusters, by site.



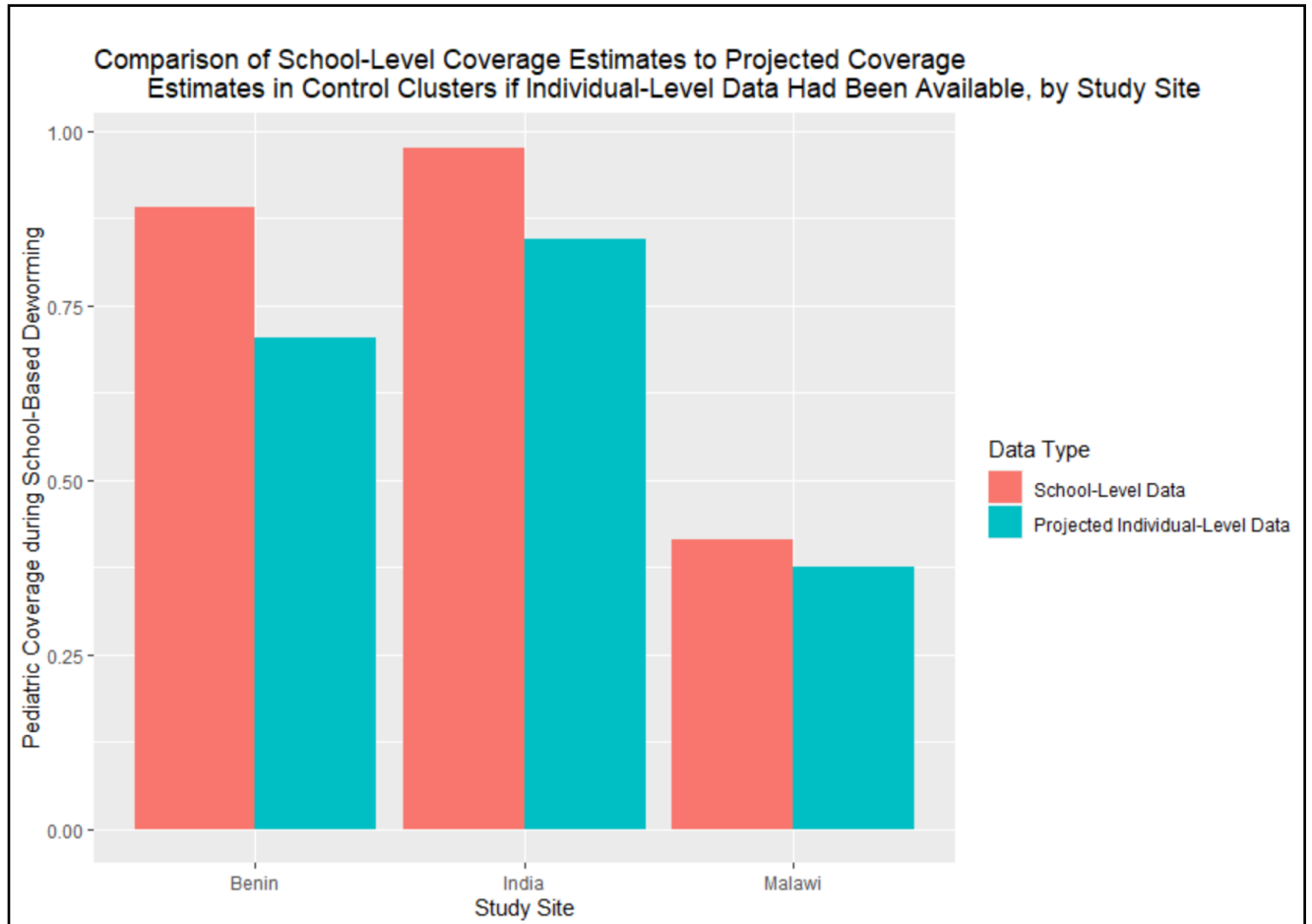
Pediatric deworming coverage from individual-level data collected through community-wide MDA in the intervention clusters was compared to school-level data collected through standard of care SBD that occurred in all clusters. By comparing the presumedly biased school-level data in the intervention clusters to the gold standard individual-level data in those same clusters, we calculated a mean squared error (MSE) for each cluster which were then averaged to determine a single error value for each site. These errors varied from 0.096 in Malawi to 0.209 in Benin ([Figure 5](#)). This difference in MSE reflects the heterogeneity in bias seen in school-level reporting between study sites. In each site, SBD coverage was reported as being substantially higher in the school-level datasets than in the gold standard individual-level datasets, which indicates that school-level SBD reporting may be over-estimating pediatric deworming coverage when compared to a gold standard of individual-level reporting.

**Figure 6:** 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.



The overall mean squared error measurements per site were applied to school-based coverage estimates in the control clusters to estimate what coverage rates might have been reported if individual-level data collection had been available in the control arm of the DeWorm3 Trial. In Benin, school-level drug distribution data from the Trial's control clusters originally showed an average pediatric coverage measurement of 89.1%, compared to 70.5% when applying observed bias to project expected coverages if individual-level data collection had been available. In India, the expected coverages for control clusters decreased from 97.7% to 84.5%, and in Malawi, projected coverages decreased from 41.5% to 37.5% (Figure 7).

Figure 7: Comparison of school-level measurements of average SBD coverage in control clusters to average SBD coverage in control clusters after application of mean-squared error rates.



## **Discussion**

The success of global soil-transmitted helminth 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 are able to 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 levels of bias inherent in school-level data collection when compared to a gold standard of individual-level data collection. This paper compares SBD data collected via individual-level MDA from the DeWorm3 Trial's intervention arm against SBD data collected via school-level reporting in both arms in order to calculate the difference between these estimates in the intervention arm clusters where both numbers are available. This is a new approach for determining how bias in data collection may result in over-estimation of pediatric 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.

SBD coverage levels for SAC derived from individual-level data collection were substantially lower than SBD estimates in the same clusters derived from school-level reporting. When applying the mean squared error between these estimates to coverage levels from the control arm, we are able to show that school-level reporting of SBD may be overestimating pediatric deworming coverage at the cluster level. This is an important finding because school-level estimates for pediatric 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 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 the premature cessation of MDA or other disruptive allocation of resources based on a flawed assumption of treatment coverage.

There could be a number of reasons for the difference 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 or PSAC who showed up for SBD days that do not usually attend that school. 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 Malawi where there were fewer students listed as enrolled on the day of deworming than in the school census. If these SAC who are arriving for deworming treatment days but are not regularly enrolled 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 ([12](#), [13](#), [14](#)), and when comparing national reporting of school-based treatment to WHO versus individual-level Demographic and Health Survey results ([15](#)). 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 ([5](#)). 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

ink to mark the fingers of SAC who were treated at school in order to mitigate this concern (16). 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 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 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 pediatric deworming treatment coverage. Previous approaches have relied upon proximity-based (17), bayesian modeling (18), and cost-distance approaches (19) to estimate likely catchment areas for service delivery 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 (20, 5) has been previously identified as a possible driver of variation in deworming treatment coverage. 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.

This analysis had several key strengths. The DeWorm3 Trial is an exceptionally large and well-organized endeavor in the STH space, with unusually rigorous data collection practices and available data from three different countries. Additionally, these data had 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 also be considered when interpreting these findings. First, the uneven quality of self-reported school attendance data used to assign population-based weights to each school in the cluster-level weighted average coverage calculations may have impacted these results. 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, a third 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. The largest percentage of free text entries that cannot be matched to a registered school included references to informal and nursery schools, while others refer to schools outside of the study area. As an illustrative example, it is believed that over 97% of the unmatched free text entries for the India study site of the DeWorm3 trial refer to Anganwadi centers. This has been a limitation in previous studies, including in Bangladesh, where informal schooling complicated calculation of coverage and population statistics ([21](#)). 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. It is also possible that a small number of free text entries were misallocated to the wrong schools in the school registry. This would only provide a significant limitation in the event that large numbers of SAC were assumed to attend an erroneous school, leading to a slight overweighting of that school in the weighted average coverage formula. Finally, it is important to note that not all available data were used. This analysis specifically utilized data collected in year two of the DeWorm3 Trial, as these data were determined by DeWorm3 Program Staff

to be the most complete dataset available. In year one, data were somewhat 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 in order to determine if pediatric coverage rates fluctuated significantly from year to year, and if observed levels of bias in the school-level SBD data were sustained throughout the duration of the trial.

## **Conclusion**

This analysis utilized data from the DeWorm3 Trial to quantify discrepancies between cluster-level estimates of pediatric 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 indicate that school-level reporting of pediatric deworming coverage likely overestimate program reach and effectiveness for school-based deworming. These findings suggest that current estimates of coverage derived from school-based program data may substantially overestimate true pediatric coverage. Novel interventions to improve data collection within MDA programs are needed to ensure that accurate reporting informs programmatic decision making and resource allocation.

## 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.
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](https://www.ntdsupport.org/sites/default/files/uploads/docs/resources/Coverage%20Evaluation%20Guidelines%20Final%20Draft_Nov%202016.pdf)
4. 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. 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.
6. 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.
7. 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.
8. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
9. QGIS Development Team, 2021. QGIS Geographic Information System. Open Source Geospatial Foundation. URL <http://qgis.org>
10. Zheng - Methods of Evaluating Estimators.pdf [Internet]. [cited 2021 Jun 10]. Available from: <http://people.missouristate.edu/songfengzheng/teaching/mth541/lecture%20notes/evaluation.pdf>
11. 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/>
12. Katarbarwa 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.
13. 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.
14. 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.
15. Lo NC, Gupta R, Addiss DG, Bendavid E, Heft-Neal S, Mikhailov A, et al. Comparison of World Health Organization and Demographic and Health Surveys data to estimate sub-national deworming coverage in pre-school aged children. *PLOS Neglected Tropical Diseases*. 2020 Aug 17;14(8):e0008551.

16. 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.
17. Cattaneo A, Nelson A, McMenemy 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>
18. 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.
19. 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.
20. 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.
21. 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.