Data Validation for School-Based ITN Distributions: Lessons Learned from Tanzania, Ghana, and Nigeria

October 2016

Angela Acosta, Claire Gillum, and Sean Blaufuss
VectorWorks Project
Johns Hopkins Center for Communication Programs

USAID FROM THE AMERICAN PEOPLE
CDC
U.S. President's Malaria Initiative

VECTORWORKS
Scaling Up Vector Control for Malaria Prevention

Acknowledgements

Many people contributed to this report. Prince Owusu (Johns Hopkins Center for Communication Programs [CCP] Ghana), Felix Nyanor-Fosu (CCP Ghana), David Dadi (CCP Tanzania), and Emmanuel Obi (Tropical Health Nigeria) led the data quality assurance activities in their respective countries. They graciously agreed to be interviewed for this report. We also appreciate the helpful feedback from Matthew Lynch, Hannah Koenker and Ato Selby.

This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID) and the President's Malaria Initiative under the terms of Cooperative Agreement AID-OAA-A-14-00057. The contents do not necessarily reflect the views of USAID, PMI, or the United States government.

Cover image credit: © Sarah Hoibak/VectorWorks 2016.

Suggested citation: Angela Acosta, Claire Gillum, and Sean Blaufuss. *Data Validation for School-Based ITN Distributions: Lessons Learned from Tanzania, Ghana, and Nigeria*. VectorWorks Project, Johns Hopkins Bloomberg School of Public Health Center for Communication Programs. 2016.

Contents

Abbreviations

CCP	Center for Communication Programs
EMIS	Education Management Information System
ITN	insecticide-treated net
LGA	local government area
MOE	Ministry of Education
MOH	Ministry of Health
NMCP	National Malaria Control Programs
PO-RALG	President's Office for Regional and Local Government
RDQA	Routine Data Quality Assessment
SOP	standard operating procedure
USAID	United States Agency for International Development
WEC	Ward Education Coordinator

Summary

In the past five years, National Malaria Control Programs (NMCPs) in at least six countries (Ghana, Nigeria, Senegal, Tanzania, Zambia, and Zimbabwe have distributed insecticide-treated nets (ITNs) through schools. The scale of these programs has ranged from a few districts or local government areas (LGAs) to nationwide. All of the programs relied on school enrollment data to determine how many nets should be allocated to storage points and distribution sites. For this reason, it is important to ensure that processes are in place to validate school enrollment data.

One of the advantages of school-based ITN distribution is that programs can use existing school enrollment data to determine the number of students who will receive nets, rather than visiting each household. School enrollment data is compiled and submitted from the classroom level to each successive administrative unit until it reaches the national level.

Because there are far fewer schools than households, and because school staff generally have higher literacy and numeracy skills than community volunteers, it is more feasible for NMCPs and implementing partners to manage this data collection process than to conduct house-to-house registration. However, school enrollment data can sometimes be inaccurate. Enrollment and attendance fluctuates throughout the year and manual data collection is subject to human error. Moreover, much of the data that are readily available at the national level are usually out-of-date by several months to a year.

This report is written for those who design and implement school distribution programs. First, it describes best practices for ensuring that enrollment data is complete and accurate. Then it outlines how programs in three of the six programs above - Tanzania, Ghana, and Nigeria - are validating the quality of school enrollment data. Other resources, such as a checklist for reviewing school distribution guidelines, an illustrative training exercise, and a validation tool are also provided.

Background

The Johns Hopkins Center for Communications Programs' VectorWorks project and its predecessor, the NetWorks Project, are President's Malaria Initiative (PMI)-funded projects that have supported ITN distributions in several countries. The school-based ITN distributions in Nigeria and Ghana were supported by NetWorks, while the 2015 school-based distribution in Tanzania was supported by VectorWorks. Earlier school distributions in Tanzania were led by the Red Cross (2013) and Research Triangle Institute (RTI) (2014).

Validation processes emerged from experiences in the first two years of school distribution in Tanzania. To minimize costs, the Red Cross and RTI were advised to use enrollment data that was previously collected by the Ministry of Education (MOE). Unfortunately, there were many discrepancies in these data sets, resulting in some schools receiving too few nets and others receiving too many. To address this issue, RTI decided to use "roving warehouses," where trucks were dispatched to each school with slightly higher numbers of nets than anticipated, with the final allocation being determined when trucks arrived at the school and could verify the actual numbers of students present. When VectorWorks took the reins in 2015, PMI Tanzania asked the project to improve the accuracy of class data. The project decided to require class teachers to submit up-to-date enrollment data two months before issuing. VectorWorks introduced validation processes to check the quality of the data prior to transporting nets. It was hoped that these steps would reduce transport costs, increase transparency and minimize the need to monitor and redistribute leftover nets. Since this experience, VectorWorks has recommended that school distribution programs validate their enrollment data before transporting and distributing nets.

Validation processes vary slightly by country but they follow a general procedure. First, enrollment data is requested by the NMCP. Class teachers submit lists of students to head teachers, who add up the numbers for the school. Data from school enrollment is then submitted to the district where it is compiled, then sent to the region, and subsequently, to the national offices of the MOE. Once all the data has been collected, the data set is compared to an estimate or to previous years' data. If the difference between the two numbers exceeds a predetermined range, then the school is flagged for a follow-up visit or phone call. After the issue has been resolved, the number of nets needed for each class, school, and district are entered into a micro-plan. Micro-plans are detailed transport, storage, quantification, and supervision plans for districts. They list the numbers of ITNs that will be transported to each level, the estimated distances between transport points, the costs for fuel and supervision allowances, and the numbers of personnel needed, and other crucial details. Once they are in place, ITN transport and distribution can begin.

The Value of Accurate Enrollment Data

Nets are bulkier and more expensive to transport than medicines, so ITN distribution programs place a high premium on accurate quantification. Significant over- and under-enrollment can mean that excess nets need to be transported or "repositioned" between schools or between districts, resulting in additional costs and management burden. Accurate enrollment data enables programs to use scarce resources efficiently. It helps ensure that the correct amount of nets, vehicles, storage, and time are allocated to the program, and thus, keep logistics costs at a minimum. The desire to reduce the costs of repositioning nets was one of the main reasons for introducing validation visits in Ghana and Tanzania.

Having a well-documented and transparent process for collecting and assuring the quality of school enrollment data is also an important part of demonstrating accountability. Donors, implementing partners, and government staff will find it difficult to be confident that intended beneficiaries receive the nets if the quantification process is not robust.

Several practices can improve data quality. These include validation visits, where supervisors conduct in-person spot checks in schools with significant discrepancies in enrollment data. Other practices include hands-on trainings, adopting user-friendly and automated forms, and keeping higher-level authorities informed and involved.

Validation visits are particularly valuable when school-based ITN distribution is new to an area. These visits show staff that enrollment data is actually being used (not just filed away), their work is being observed, and that data quality matters. Lastly, validation visits provide a platform for correcting misunderstandings and preventing costly mistakes.

Characteristics of Data Quality

School enrollment data are considered good quality if they meet the following criteria:

1. **Complete:** All eligible schools and classes submitted data, not just some or most, and all eligible students in eligible classes were registered. The definition of eligible classes can vary by country. In Ghana, for example, the program determined that eligible classes were Primary 2 and Primary 6; therefore, it sought data only from those classes. Similarly, in Nigeria, the program requested data only for Grade 1, Grade 4, Junior Secondary School 1, and Senior Secondary School 1. In Tanzania, the program requested enrollment data for all classes during the third round of the school net program, and then selected the eligible classes.
2. **Timely:** Data advances to the next administrative level in time for data quality checks. The process of data submission and finalization did not cause distribution delays.
3. **Accurate:** The data contained minimal recording, summarizing, or transcription errors.
4. **Reliable:** Similar processes were used for the collection and dissemination of enrollment data, to ensure that data were collected consistently
5. **Integrity:** The data was protected from deliberate bias or manipulation for personal or political reasons.

Common Causes of Errors

The following issues can lead to significant under- or over-reporting:

1. **Missing data:** When data were compiled at the national level in Tanzania, program officials discovered that 103 of the 1,919 schools (5.3%) were missing data for at least one class.
2. **Misinformation:** Most of the inaccurate data in Nigeria was due to a misunderstanding among school heads that each ITN would incur a cost to the school. This led to intentional under-reporting. A cascade training strategy was probably the source of this misinformation. After understanding that the nets were free, school heads submitted corrected reports. Similarly, there can be misunderstandings about who is and is not eligible for nets. Are students in private schools, boarding schools or in special needs classes eligible for nets? What about teachers? These questions have come up in Tanzania and other countries.
3. **Numbers are not correctly recorded onto new forms:** At each level, many numbers are incorrectly recorded onto new forms. For example, in Tanzania, the number of girls and boys for all six eligible classes have to be recorded onto the school summary forms, then the district summary forms. In the process, numbers can be transposed or incorrectly copied. These mistakes are then incorporated into the regional, and ultimately, national summaries.
4. **Incorrect arithmetic:** Many of the forms require school and health personnel to add the total numbers of students before submitting them to the next level. Sometimes, several rounds of additions need to be made to complete one form. Some forms require totals for each row, then for each column. The more arithmetic involved in filling out forms, the more likely it is for errors to occur.
5. **Out-of-date data**: During the 2014 school distribution in Ghana, the NetWorks project relied on data from the Education Management Information System (EMIS), which collects enrollment numbers and other information from an annual survey of schools. However, the EMIS excluded many private schools and the data was often not available until a year after it was collected, making it outdated by the time it was available for planning a school distribution.

Recommendations

Because school-based ITN distributions take place once a year, they share many similarities with mass campaigns. In fact, many of the recommended processes for managing school enrollment data were adapted from mass campaigns.

1. **Ensure that there is ample time for compilation, data quality assurance, and validation.** The process of compiling data at the school level for quantification at the national level can take several months. Delays in finalizing the enrollment figures can lead to postponement to school distribution, which can cause increased storage costs and unnecessary management burden. In Tanzania, delays in the 2015 submission of enrollment data caused ITNs to remain at the national-level warehouse for longer than planned. This led to transportation companies working extended hours to ensure that trucks departed for the districts before the storage contract expired. To avoid these delays, program planners should set aside enough time to ensure that data is compiled and checked at each level and conduct validation visits.

Each program will need to find a balance between starting early enough for these processes to happen, but starting late enough in the school year that class enrollment will have stabilized. Typically, programs start collecting data two to five months prior to distribution. To prepare for the 2016 distribution, Ghana requested enrollment data at the beginning of the first term, with hopes that the final data set would be ready midway through the second term. This gave the various levels (circuit, district, and region) 16 weeks to compile, check, and submit their data, 3 weeks for national-level compilation, and 2 weeks for validation, before issuing ITNs in the third term.

It is vital to involve school officials in developing this timeline. Similarly, it is important to communicate this timeline well in advance. Staff may need to meet multiple deadlines simultaneously. They are more likely to submit complete and accurate reports in a timely way when they can anticipate their workflow.

2. **Every person submitting data should check for errors before submission.**
In most countries, school enrollment data flows up to the national level through the next immediate level of authority. This creates an opportunity for checks as well as errors. Thus, class teachers should check their data before submitting it to head teachers or principals. Similarly, head teachers should check their tallies before submitting it to subdistrict supervisors (such as Ward Education Coordinators in Tanzania or Circuit Supervisors in Ghana). These supervisors should review the data before submitting to district school planners, and the same should be done at the regional and national level.

   Many supervisors make phone calls or conduct spot checks in person to verify the data they receive. Trainings and coordination meetings should highlight examples of validating data and encourage them as good practice. To encourage consistency, these training and coordination meetings should include participants at all levels of the quantification process.

   To further ensure these responsibilities are well-defined, quantification guidelines should describe:
   - The process to follow when an error is discovered in a report;
   - Who is responsible for notifying the school if errors are discovered;
   - Who conducts the initial follow-up on the initial notice of a problem; and
   - What tools or other assistance are available to schools as they try to correct errors.

3. **Trainings should include hands-on exercises in checking and fixing forms.**
Tanzania, Ghana, and Nigeria have typically used cascade training systems. A core team of implementers, NMCP, and MOE staff train regional and district trainers, who then train subdistrict supervisors or schools. Participants currently receive training on how to fill out the forms and they are asked to check the forms before submitting them to the next level. However, participants should also conduct hands-on practice in checking and fixing incorrectly completed forms. Incorrectly submitted forms from previous school distribution rounds can be adapted for training exercises. Where appropriate, some training in troubleshooting with Microsoft Excel spreadsheets should be considered.

   The importance of good quality trainings cannot be emphasized enough. When new cadres of personnel are asked to participate in school distributions or when procedures have been significantly revised, it is important to consider how trainings will be carried out. Trainings currently take no more

than a day, and it may be hard to pay for longer trainings at scale. However, having sufficient copies of clear and user-focused training materials, exercises, and job aids can go a long way in making the most use of participants' limited time.

In 2015, Tanzania involved the Ward Education Coordinators (WECs) in the School Net Program for the first time. Many significant changes were also made in the program's design. During a procedural audit, WECs reported that the one-day training was not enough to learn all of the content. They also found that the standard operating procedures (SOPs) were confusing, because they did not clearly delineate who was responsible for carrying out the procedures.

Validation visits and supervision of issuing also showed that many WECs did not adequately understand elements of the SOP. Supervision teams encountered several instances in which WECs had misunderstood procedures to be followed, such as failing to distribute data collection tools at the appropriate time or not conveying accurate information about which classes were eligible to receive ITNs.

4. **Add data quality checks and sufficient instructions to forms. Whenever possible, forms should be automated.**

   Quantification forms should have built-in reminders. All personnel should be cued to seek and address common errors such as missing data, transpositions, and miscalculated totals before signing off on their submission.

   Digital formats such as spreadsheets, SMS, or data entry software should be considered for those at district and higher levels, where much of the aggregation occurs. Such forms can automatically flag missing numbers and add up the totals. Other data quality assurance methods such as double entry can also be used. For example, if data is entered twice, the software flags any differences and the person entering data is required to correct errors before the submission can be considered completed.

   In the cascade model of training, in which master trainers train other trainers, details can be missed during the transfer of information to the next level. At the same time, misinformation can enter the information chain. For this reason, instructions on the forms should clearly define who counts as an enrolled student. In Tanzania, for example, the SOPs state that students who are absent on issuing day should receive a net. However, it is not clear how long a student should be absent before they no longer count as enrolled for the purpose of the school distribution program. This is an issue because many students are often absent for weeks at a time. Moreover, in Ghana, some urban districts had insufficient nets during the 2014 school distribution because the enrollment figures did not factor in private schools who were not registered with the Ghana Education Service. Clarifying these issues in the enrollment and compilation forms could have prevented unnecessary over- and under-reporting.

   Carbon paper should be attached to forms so that each submitting level can keep a copy. These will be useful if someone calls to verify data quality or conducts spot checks during validation visits; they can also be used to verify the numbers of nets allocated by the Ministry of Health (MOH) during process or post-distribution evaluations, or supply chain audits.

5. Conduct validation checks and validation visits to resolve significant remaining discrepancies

Enrollment data should be compared to a second data set once it reaches the national level. This second data set should enumerate the same target population. For example, in Tanzania's third round of the school net distribution pilot, *current enrollment data* using school distribution forms were compared to *issuing data* from the previous two rounds. Program officials compared the number of students reported as currently enrolled in Class 5 to the number of Class 4 students who received nets last year.

During the 2014 school distribution in Ghana, the NetWorks project relied on data from the Education Management Information System (EMIS), which collects enrollment numbers and other information from an annual survey of schools. However, the EMIS excluded many private schools and the data was often not available until a year after it was collected, making it outdated by the time it was available for planning a school distribution. In 2016, Ghana is planning to request *current enrollment* data from schools (collected using forms expressly designed for this purpose) and compare that data to the number of students enrolled in the 2015 EMIS; in this case, Primary 6 students will be compared to the number of Primary 5 students.

Next, a variance threshold should be selected. Schools with enrollment numbers that are a certain number of percentage points above or below the previous year's numbers should be flagged for validation. In Tanzania, the MOE and MOH selected 30% as the threshold, because this was the percentage of students who started but did not complete primary school. However, this was based on the *national* intra-school survival rate of 70%. Given that some areas have higher and lower survival rates compared to the national aggregate, the team should use regional or district information (if available) to determine a more precise threshold. Thresholds can vary from country to country, so real data should be used whenever possible. In Ghana, for example, the threshold for the 2016 school distribution was 15% as of the writing of this report.

Whenever possible, subnational survival rates should be obtained to make these thresholds more likely to reflect true incongruities. A variance threshold must be carefully selected to balance concerns of sensitivity and specificity. If the threshold is too low and, therefore, identifies a large number of sites for validation, the activity may become costly and unsustainable at scale. However, if the threshold is too high and sites with real data quality issues are not flagged, the validation activity will be unlikely to achieve its objectives.

In Tanzania, current enrollment numbers and comparison data were entered into a Microsoft Excel spreadsheet. The spreadsheet was formatted to include conditional formatting rules, causing the font color of the current enrollment number to turn red when it is below or above the agreed-upon threshold compared to the comparison data set. When spreadsheets are used, enrollment and subnational survival rates can be linked using the pivot tables feature.

When comparison data is not available, a random sample of schools can be selected for validation visits. The sample can be stratified by region, district, school size, whether it is public or private, and urban or rural.

Conduct validation visits. Once schools have been flagged, teams should be dispatched to validate the data. Such visits may include the following:

- Cross-checking quantification data against class registers at schools, and against copies of summaries at the subdistrict, district, and regional levels.
- Checking the calculation and aggregation of data across the various forms each level of reporting.
- Sampling classes or schools to confirm the number of enrolled students reported.
- Meeting with local authorities to confirm that the procedures were followed as planned.

Update the data set and micro-plans. After the quality checks and validation visits have been completed, the enrollment data should be updated and shared with all reporting levels (including the schools) and with all relevant stakeholders. This data is then used to inform the allocations and micro-plans.

6. **Review the data as a group before finalizing the micro-plans.**
   During mass campaigns, district officials meet at the regional level to review the data and finalize the micro-plan. During this meeting, participants check the data for errors and make phone calls to obtain missing or corrected information.

   School distribution programs can also review enrollment data during micro-planning meetings. Costs can be kept to a minimum if validation visits and micro-planning meetings take place immediately before or after each other.

7. **Have a plan for dealing with variances discovered during the issuing phase**
   In Ghana, during both the 2014 and 2016 school distributions, circuit supervisors were given the ability to fix enrollment forms during the issuing phase. When they transported nets to schools, they checked the enrollment numbers and adjusted the forms. This enabled them to issue the exact number of needed nets to schools. In this system, the corrected enrollment forms had the same totals as the distribution forms.

   However, during Ghana's 2014 school distribution, leftover nets stayed in the district stores, with no plans to distribute them before the 2016 distribution. These nets sat unused in dispersed locations for two years, creating opportunities for theft and damage. Wherever possible, leftover nets should be formally handed over to the routine health distribution system (with waybills to document the transfer from district education stores to district health stores, where appropriate). Ghana and Tanzania both adopted this procedure in 2016.

   Decisions such as how to document variances and how to handle leftover nets (and who will pay for it) should be discussed during the planning meetings held at national, regional and district levels with the relevant coordinating committee (this usually comprises of members of the ITN technical working group, MOE and NMCP (see Roles and Responsibilities). Transporting the small amount of nets left over from schools to places such as health facilities (as was done in Tanzania and Zimbabwe in 2015) or to the district stores (Nigeria 2014) has usually been incorporated in supervision visits and have thus cost little.

Roles and Responsibilities

Roles and responsibilities for data quality checks should be agreed upon early in the planning and advocacy process. Data quality checks can be tedious, especially for district and subdistrict supervisors, who collect and tally much of the raw data. Thus, it is very important to strengthen their belief that these are important procedures to undertake. Concurrently, programs should strengthen the capacity of supervisors during trainings, provide them with job aids and easy-to-use forms, conduct feedback sessions, and ensure that their supervisors (the regional authorities) strongly support the program and are up to date on the proceedings.

School distributions usually are implemented jointly by the MOE and MOH. Within the MOE, the focal point at the national level is the coordinator of the School Health and Nutrition Program (also known as the School Health Education Program in Ghana), while the focal point in the MOH is the ITN focal person of the NMCP. NMCPs have usually conducted school distributions with the assistance of an implementing partner such as VectorWorks.

In most of the countries, both ministries (MOE and MOH) develop and conduct joint trainings for staff. ITNs are health commodities and may thus require slightly different calculations compared to educational commodities. Training materials should be developed by MOH and finalized after MOE review.

Data is typically submitted through the MOE's chain of command. Thus, responsibilities for checking data at the regional level and below will typically fall under MOE. At the national level, data is handed over from the MOE to the MOH's NMCP. At this point, the NMCP is responsible for conducting a final check of the data and conducting validation visits; if available, in-country partners may assist as well. Finally, the NMCP shares the ITN allocations with MOE.

Each country has a unique group of lead implementers for school distribution. In addition to VectorWorks, NMCP, and MOE, the President's Office for Regional Administration and Local Government (PO-RALG) also played an important role in the data quality components of the third round of the school distribution program in Tanzania. While the other two ministries provided technical oversight, the PO-RALG had executive oversight (budgetary and hiring authority) and therefore generally received high compliance from all ranks. The PO-RALG representatives for education and health participated in all planning meetings. When significant questions of data quality emerged, the regional representatives of this ministry (Regional Administrative Secretaries) instructed the relevant subdistrict supervisors to resubmit their data. Similarly, the Nigeria State Education Board that checks classroom registration data has direct authority over school heads; therefore, engaging them in the data quality process gets an immediate response from schools heads. By keeping those with executive authority informed, the Tanzania and Nigeria programs were able to significantly increase accountability for data quality.

Involving these key stakeholders during advocacy, micro-planning, and training has been helpful in getting them to understand the importance of accurate enrollment data. Those who buy into the program can send a strong message to other staff and help build a culture of data quality.

Resources Needed

The costs of incorporating data quality checks can be minimal and the resulting savings can well offset any expenses incurred. Many of the recommendations above, such as checking data before submission, strengthening forms and training materials, and reviewing data during micro-planning meetings, cost relatively little. However, they can greatly improve the quality of the data, minimize logistics costs, and significantly strengthen confidence in the quality of the school distribution program among government staff and donors. In Tanzania, for example, validation revealed there were 9,586 fewer eligible students than originally reported, which represented 1.9% of the total ITNs available for distribution.

The checks mentioned above can also minimize the costs of validation field visits, as they should greatly reduce the number of schools that will need to be visited. In Tanzania, for example, 335 (17%) of all participating schools were flagged for validation visits; of these, one-third had been flagged due to missing data. Validation visits occurred over a period of 20 days and involved approximately nine personnel from VectorWorks, MOH, and PO-RALG at the national level, in addition to regional and district-level government representatives. The total cost of fuel, transport, and per diem (not including salaries) was approximately $15,500. When one considers that a third of the visits could have been prevented, it is fair to say that the cost and management burden of validation could have been significantly reduced if the requisite steps had been followed. Carrying out such large numbers of validation visits could be a hindrance to implementing school distribution at scale. Ideally, validation visits would be required for less than 5% of schools (about the same rule of thumb used for spot checks). Reducing the percentage of schools that require validation visits can be achieved by strengthening data management processes, trainings and forms (see Recommendations). If/ when validation tools reveal that more visits are required than is feasible (given time and money), program planners will need to decide how many to visit, how to sample from the identified schools, and how to address widespread corrections that are identified during these validation visits.

National programs typically bear the bulk of the cost of validation visits. To keep costs and time requirements manageable, some countries such as Tanzania have chosen to use a large variance (such as 30%) between reported and expected number of students as a justification for conducting validation visits. As government and school staff strengthen their capacity in data quality assurance, data quality will improve and the number of validation visits and related costs will likely decrease.

Electronic systems, such as the submission of data through fixed spreadsheet templates (Annex 4), SMS surveys or online portals, could also reduce the burden of validation on staff. These systems can automatically notify those entering data when the submitted numbers are out of range, in an invalid format, or when data is missing. The costs of developing such systems can be offset when open-source software is used.

Conclusion

School distribution depends on enrollment data in order to allocate nets. Having good-quality data can help minimize the costs of logistics and strengthen donor and government confidence in the program. This report describes several ways to bolster the quality of this data. Ultimately, the processes for validating school registration data will vary from country to country, depending on factors such as the availability of comparison data, variability in enrollment throughout the school year, capacity of government staff at various levels, and other factors. For these same reasons, the final numbers used in the program will never be perfectly accurate. However, each program should strive to achieve a balance between having reliable and complete data and being able to implement ITN distributions in a timely manner that is not unduly burdensome. Thus, it is important for programs to do their due diligence and incorporate data quality assurance processes in school distribution guidelines.

Annex 1. Country Experiences

Table 1 compares data validation processes in school-based ITN distribution programs in Tanzania, Ghana, and Nigeria. Each undertook slightly different methods to register eligible students and verify data quality. It should be noted, however, that only Tanzania has completed a school-based distribution with formal validation activities. Ghana will begin a formalized validation process prior to the 2016 school-based distribution, and validation visits were done ad hoc in Nigeria.

Despite the differences highlighted in Table 1, several lessons emerged that should be considered for future school-based ITN distributions. Cascade trainings are practical, but can lead to misinformation and confusion, particularly at lower administrative levels where implementation occurs. In Nigeria, underreporting resulted from lack of understanding among school heads, and in Tanzania, the training was cited as a primary cause of missing data. Changes in training methods or training materials may be useful to improving initial quantification data.

Monitoring and data checks at every level of aggregation increases the likelihood of resolving inaccurate data. Major sources of inaccurate data in Tanzania were improper data collection processes compounded by insufficient cross-checking of data during the aggregation process. Procedures that increase accountability for following protocol will likely improve initial quantification at subdistrict levels, where most data quality issues arise.

Validation of enrollment data is an important consideration for any ITN distribution mechanism; however, it must be done in a cost-effective manner. Formal validation activities in Tanzania were informative, but resource intensive. No information is available to confirm the effectiveness of the data quality assurance system in place in Nigeria, but validation was done at considerably less cost. Improving trainings, forms, and job aids, and building in checks at each level of submission will nurture a culture of data quality and reduce the cost of validation over time.

Table 1: Data validation processes in Tanzania, Ghana, and Nigeria, 2012–2015

Processes	Tanzania (2015)	Ghana (2014)	Nigeria (2012–2014)
Data source	Schools submitted data just for the school distribution program.	Used existing EMIS data.	Schools submitted data just for the school distribution program.
Data flow	School teachers → head teachers → WECs (ward level) → district → regional → national	National level only. MOE (School Health Education Program Unit) → NMCP → VectorWorks Since available data is one year old, they used the number of students from P1 and 5, instead of P2 and 6.	School teachers → school head → LGA → State Education Office → MOH
Data quality checks/ adjustments	At the national level, current school enrollment data was compared to class cohort data from the previous two years. If the variance exceeded 30% or if class data was incomplete, the school was flagged for validation.	A 10% buffer was added to enrollment data and this was the number of nets allocated to districts. Head teachers went to the district to pick up the number of nets they needed based on current enrollment (the names of students were submitted as proof).	After data collection, each higher aggregation level examines data based on historical enrollment and judgment. During the micro-planning stage, all stakeholders agreed on enrollment data.
Validation actions	Schools with greater than 30% variance in enrollment from previous School Net Programs were subject to field visits. Four to six national and regional personnel cross-checked enrollment data by hand, and questioned school and national authorities.	There were no formal validation checks.[a]	There were no formal validation checks. If poor data was suspected, the school headmaster was contacted. The State Education Board had the authority to hold school heads accountable if necessary.
Timeline	Schools submitted data in early May 2015 and submitted it to the national level at the end of May. Validation visits were conducted over a three-week period in July.	EMIS data was received and compiled for distribution in November 2013.	Data was collected during the second term (when enrollment was most stable) using classroom register data. Data was requested three days before the micro-planning/training meeting.
Issues faced	5.3% of submitting schools had incomplete data, and another 12% had variances greater than 30%. The validated data from the	Most rural districts were oversupplied while urban districts were undersupplied because the EMIS did not capture all private schools.	Some underreporting due to confusion about cost to school.

Processes	Tanzania (2015)	Ghana (2014)	Nigeria (2012–2014)
	first round was rejected; a second round of data collection was conducted. Because the distribution was already delayed, the second set of data was not validated. Some districts did not follow the allocation plans during re-bundling but used their own judgment instead.		
Plan for leftover nets	Nets were taken to a hospital in Lindi for distribution.	ITNs remained in district stores until the 2016 school distribution.	Extra ITNs were transported to the district level during supervision visits and redistributed among schools within LGAs during the monthly meeting of school heads
Lessons learned	Trainings needed to be improved to increase comprehension at lower levels. Strengthening the NMCP and implementing partners' engagement with the administrative sector may improve initial school data. Enough time must be given to compile data for validation activities to be completed before distribution.	EMIS data was not complete or timely enough to use for school distribution.	School heads must be properly trained to avoid misinformation and improper quantification

[a] The lack of validation checks will soon change. Ghana plans to introduce validation visits in the 2016 round of school distribution. Variance of greater than 15% between current enrollment figures and the EMIS will trigger an in-person validation visit to check data, and ensure that it is corrected or authenticated.

Annex 2. Data Quality Assurance Checklist

This checklist was adapted from MEASURE Evaluation's Routine Data Quality Assessment Tool (RDQA). It can be used to assess gaps in the processes used/planned for enrollment data.

Table 2: Data quality assurance checklist

√ or N/A		I. Structure, Functions, and Capabilities
	1	MOH and MOE designated staff at each level (national, regional/province, district, subdistrict) who are responsible for reviewing the quality of data (i.e., accuracy, completeness, and timeliness) received from sub-reporting levels. This can include one person from each ministry (MOE and MOH) or a focal person jointly selected by the participating ministries.
	2	MOH and MOE designated staff at each level (regional/province, district, subdistrict, school) who are responsible for reviewing aggregated numbers prior to submission to the next level. This can be the same person(s) identified in step 1
	3	All relevant MOH and MOE staff have received training on the data management processes and tools.
II. Indicator Definitions and Reporting Guidelines *Written guidelines are available to regional/provincial, district, subdistrict and school levels on …*		
	4	*What* they are supposed to report on.
	5	*How* (e.g., in what specific format) reports are to be submitted.
	6	*To whom* the reports should be submitted.
	7	*When* the reports are due.
III. Data Collection and Reporting Forms / Tools		
	8	MOE and MOH have provided clear instructions to each reporting level (regional/provincial, district, subdistrict and school) on how to complete the data collection and reporting forms/tools.
	9	All reporting levels use standard reporting forms/tools .
	10	All *source documents* and *reporting forms* relevant for measuring the indicator(s) are available for auditing purposes (including dated print-outs in cases of computerized system).
IV. Data Management Processes		
	11	Feedback is systematically provided to all schools and on the quality of their reporting (i.e., accuracy, completeness, and timeliness).
	12	When paper forms are used, there are quality controls in place for when data are entered into a computer (e.g., double entry, post-data entry verification, etc.).
	13	When computerized systems are used, there is a written back-up procedure for when data entry or data processing is computerized.

	14	When there is a policy or procedure for backing up computerized data, the latest date of back-up is appropriate given the frequency of update of the computerized system (e.g., back-ups are weekly or monthly).
	15	Personal identifiers are stored based on national or international confidentiality guidelines.
	16	There is a written procedure to address late, incomplete, inaccurate, and missing reports; including following up with schools on data quality issues.
	17	If data discrepancies have been uncovered in reports from schools, higher aggregation Levels (e.g., districts or regions) have documented how these inconsistencies have been resolved.
V. Links with National Reporting System		
	18	Data are submitted through a single channel of the national reporting system (it can be shared with multiple partners and ministries).
	19	When available, the relevant national forms/tools are used for data collection and reporting.
	20	The system records where the nets are delivered (region/province, district, subdistrict, school, etc.)
	21	When place names are recorded, standardized naming conventions are used.

Annex 3a . Illustrative Training Exercise

This exercise is intended for subdistrict supervisor orientations in settings where paper forms are used for collecting school enrollment data. A similar format can be adapted for the district/regional orientations.

Materials: Each subdistrict supervisor will receive:

- One blank subdistrict-level form and one blank school-level form.
- Two incorrectly completed school-level forms (see Annex 3a and b for examples).
- One incorrectly completed subdistrict-level form (see Annex 3c for an example).

Steps:

1. Subdistrict supervisors will be charged with submitting a correct subdistrict-level form. To do so, they will be asked to refer to the SOPs, which will include the following checklist:

School-level forms	Subdistrict -level forms
☐ All classes are represented, with numbers for boys and girls in each class. ☐ Each class is summed correctly: o The total number of students is correct o The total number of girls is correct o The total number of boys is correct ☐ Each school is summed correctly: o The numbers from each class are copied correctly on to the school form o The total number of students is correct o The total number of girls is correct o The total number of boys is correct	☐ All schools are represented. ☐ All schools are summed up correctly: o The numbers from each school are copied correctly on to the ward form o The total number of students is correct o The total number of girls is correct o The total number of boys is correct

2. They are given 10 minutes to review the school forms for errors. After 10 minutes, the trainer asks the group to state the mistakes they found.
3. The trainer gives the subdistrict supervisors another 10 minutes to review the subdistrict-level forms, whereupon the necessary corrections are discussed again.

Annex 3b. Sample School Form 1

Instructions: *This form will be used by subdistrict coordinator to collect data in all schools in his/her ward. A copy of this form should be given to the head teacher; subdistrict coordinator should retain one copy and copy submitted to the district.*

Name of Region: ABC **Name of District:** Mbinga

Name of Subdistrict: Upolo Name **of Village**: Lunhido

Name of School: Chomba Primary

S/N	Classes	Gender		Total
		Male	Female	
1	Grade 1	21	18	39
2	Grade 2	8	13	21
3	Grade 3	23	30	53
4	Grade 4	32	24	56
5	Grade 5	127	148	275
6	Grade 6	18		36
7	Grade 7	47	37	84
	School's TOTAL	**276**	**288**	**564**

Recommendation/remarks (if any)

_____

_____

_____

___

NB: *This report will be valid and useful if and only if it will be signed and stamped by both subdistrict coordinator and Head teacher*

Name of Head Teacher_____

Signature of Head teacher_____ **Date**: _____

Name of subdistrict coordinator_____

Signature of subdistrict coordinator _____ **Date**: _____

Annex 3b. Sample School Form 2

Instructions: *This form will be used by subdistrict coordinator to collect data in all schools in his/her ward. A copy of this form should be given to the head teacher; subdistrict coordinator should retain one copy and copy submitted to the district.*

Name of Region: ___ABC_____ **Name of District:** _____Mbinga_____

Name of Subdistrict: ___Upolo_____ Name **of Village**: _____Lunhido_____

Name of School: Mihigo Primary_____

S/N	Classes	Gender		Total
		Male	Female	
1	School 1	35	27	62
2	School 2	12	20	32
3	School 3	28	40	68
4	School 4	37	25	62
5	School 5	145	135	260
6	School 6	27	36	63
7	School 7	56	48	104
	School's TOTAL	**304**	**331**	**651**

Recommendation/remarks (if any)

_____

_____

_____

___

NB: *This report will be valid and useful if and only if it will be signed and stamped by both subdistrict coordinator and Head teacher*

Name of Head Teacher_____

Signature of Head teacher_____ **Date**: _____

Name of subdistrict coordinator_____

Signature of subdistrict coordinator _____ **Date**: _____

Annex 3c. Sample Subdistrict Form

Instructions: *This form will be used by subdistrict coordinator to collect data in all schools in his/her ward. A copy of this form should be given to the head teacher; subdistrict coordinator should retain one copy and copy submitted to the district.*

Name of Region: ___ABC_____**Name of District:** _____Mbinga_____

Name of Subdistrict: ___Upolo_____

S/N	Classes	Gender		Total
		Male	Female	
1	Berebere Primary	420	385	805
2	Obenga Primary	144	240	384
3	Upolo Primary	336	480	816
4	Essaye Primary	444	300	744
5	Chomba Primary	276	288	564
6	Mihigo Primary	331	304	651
7	Garuda Primary	672	576	1248
	School's TOTAL	2623	2573	5212

Recommendation/remarks (if any)

_____

_____

_____

___

NB: *This report will be valid and useful if and only if it will be signed and stamped by both subdistrict coordinator and Head teacher*

Name of subdistrict coordinator_____

Signature of subdistrict coordinator _____ **Date**: _____

Annex 3d. Answer key

School form 1: Chomba Primary

1. Row Grade 6, Female: The blank field is incorrect. The sub-district coordinator should call the school to obtain the missing amount.

School form 2: Mihigo Primary

2. Column Female, Total: This total was copied incorrectly. The correct total is 340, not 304 (the last two digits were transposed).
3. Row Grade 5, Total: This was calculated incorrectly. The correct total is 280, not 260.
4. Total (bottom right corner): The correct total is 671, not 651. After correcting the Grade 5 total, the subdistrict coordinator should update the total for the subdistrict.

Subdistrict form

5. Mihigo Primary: the male and female totals are switched. However, the totals for the school and for the subdistrict are correct.

Annex 4. Validation tool from Tanzania

VectorWorks
Johns Hopkins Center for Communication Programs
111 Market Place, Suite 310
Baltimore, MD 21211
410-659-6300
www.vector-works.org