APPENDIX I – ADDITIONAL VERIFICATION EXERCISE DETAILS

The contents of this appendix provide additional details related to the initial verification exercise that may be of interest to the Foundation and/or other audience members.

1. VERIFICATION DESIGN

1.1 Identifying verification measures

For our verification, the Foundation indicated interest in verifying sanitation outcomes reported by BRAC, more specifically, the Foundation was interested in verifying latrine coverage and use. In the RFP, the Foundation indicated the specific indicators to be verified were: 1) the number of latrines constructed or repaired, and 2) the percentage of adults in households that consistently use their latrine six months after the construction/repair. The RFP also stated that the verifying entity should select a random sample of households from which monitoring data were collected by BRAC to verify monitoring data on latrine coverage and latrine usage. Since the QIS monitoring system is the only segment of BRAC's monitoring system that contains an electronic account of latrine use data, it was decided that the verification sample should be selected from within the QIS population (with a comparison sub-sample selected from any VWC touched by the project). As a result, our verification was limited in scope to the three sanitation-related indicators collected via BRAC's QIS survey. Therefore, we were not able to conduct an actual verification of some indicators set out in the RFP since they were not captured in the implementer's monitoring system. We did, however, incorporate additional indicators into our verification survey instrument which allowed us to assess the other aspects of latrine construction/repair and latrine use that were of interest to the donor (see section *1.1.2 Additional indicators assessing aspects of interest to the donor* for further details).

1.1.1 QIS indicators as key verification measures

BRAC's QIS survey instrument queries households on three sanitation-related indicators – one on the presence of a latrine (i.e., coverage with a sanitary latrine), one on whom within the household utilizes the latrine (i.e., latrine use, by whom), and one on when the household members who use the latrine actually utilize it (i.e., latrine use, when). These were, therefore, the three indicators we used as the foundation of our verification. In order to ensure comparability of QIS indicator collection between samples (BRAC's QIS sample and our verification sample), and to limit survey administration bias, we invited BRAC QIS trainers to train icddr,b enumerators on how to appropriately administer QIS survey prompts and scale household scenarios via BRAC's QIS 'ladders'.

1.1.2 Additional indicators assessing aspects of interest to the donor

The Foundation was interested in assessing wealth targeting and intervention receipt, and specified that latrine use data should be collected through: 1) survey data that captures self-reported use, 2) direct observation, and 3) and sensor data, our initial verification survey instrument collected many indicators in addition to the three key verification indicators. We incorporated additional indicators into our verification survey instrument to allow us to assess household wealth status, demographics of household latrine use, details regarding the most recent latrine construction/repair, and the timing of that construction/repair. We included other indicators in the survey instrument to allow us to assess predictors of latrine coverage and utilization (e.g., distances between the latrine and points of interest [i.e., the homestead and the nearest water source], exposure to programme "software" components, and recent illness/symptom reports [e.g., diarrhea/dysentery and respiratory symptoms indicative of ARI]).

1.1.3 Methods for measuring latrine utilization

As per the Foundation's request, we carried out three different methods for measuring latrine utilization.

1. ***Self-reported use data:*** We captured self-reported latrine use data via a structured household census use schedule. Through the use of the household use schedule, enumerators systematically captured latrine use data and defecation practices for each member of the household.
2. ***Direct observation data:*** We conducted direct observations via latrine spot checks, which are visual inspections of latrines and the nearby premises. Through the completion of latrine spot checks, enumerators systematically captured data on indicators that allowed us to directly and indirectly assess latrine use. Data were captured on the following latrine spot check indicators:
    - evidence the latrine is being used for storage (e.g., storage of household items, goods that do not belong in a latrine);
    - a well-worn path leading to the latrine;
    - a wet latrine floor;
    - odor of stool/urine in the latrine;
    - flies in the latrine;
    - discoloration of the latrine pan;
    - presence of visible traces of feces in the pan;
    - water container in/near the latrine;
    - cleaning agents inside the latrine;
    - broken pan or leaves/dirt/spider webs in the latrine pan;
    - water for hand-washing near the latrine;
    - soap/ash for hand-washing inside or near the latrine; and
    - slippers outside the latrine.
3. ***Instrument-recorded use data:*** We deployed a passive latrine use monitor (PLUM) in every functional latrine at every consenting household within a randomly selected sub-set of verification households (all households in approximately half of the QIS verification sub-sample were asked if they would agree to having a PLUM installed in their latrines) . PLUMs were deployed for a one-week installation period, and collected household level latrine use data via their electronic sensing capabilities. Sensor data were assessed for signal patterns consistent with defecation events via the use of a validated algorithm.

1.2 Selecting the verification sample

Due to the fact that the WASH II project population is widely distributed throughout Bangladesh, and because we wanted to hold constant as many variables as possible for our comparative analyses, we designed the initial verification exercise with a cluster sampling design similar to that used by BRAC/IRC to collect their QIS sample.

1.2.1 Identifying the target population

For the purposes of the original verification exercise, we initially determined that BRAC reportedly targets all "ultra-poor", "poor", and "non-poor" household within the 157 Foundation-funded WASH II project upazilas with indirect software interventions. In addition to these software interventions, BRAC reportedly targets 50% of all ultra-poor households for direct hardware intervention receipt. Therefore, our initial plan was to define the target verification population as all households situated within the 157 Foundation-funded WASH II project upazilas. We subsequently learnt that while BRAC's MIS system contained latrine utilization indicators, they were not included in the electronic MIS dataset. We therefore agreed with the Foundation that the QIS sample should serve as our study population. However, upon the receipt and

review of the actual QIS monitoring dataset, it became clear that 2832 of the 7489 observations in the dataset represented household data from the 20 WASH II-engaged upazilas supported by the Dutch Ministry of Foreign Affairs (DGIS). Therefore, we adjusted our target population definition to all households situated within the 175 WASH II programme upazilas (177 when accounting for the recent split of 2 WASH I upazilas). Adjusting the definition of the target population allowed us to maintain sufficient power to verify/detect differences in results both within and between samples at both QIS and Foundation-supported levels (i.e., allowing for inclusion of DGIS-supported upazilas/VWCs). Specifically, this sampling strategy allowed us to detect differences in effect at sub-population levels in wealth category.

1.2.2 Determining primary and secondary sampling units

In order to obtain their QIS sample, BRAC reported in their QIS manual that they sampled equally from WASH I-engaged communities and WASH II-engaged communities. Since WASH I was comprised of 152 upazilas (i.e., sub-districts), and WASH II was comprised of only an additional 5 upazilas supported by the Foundation and 20 upazilas supported by DGIS, BRAC selected two different primary sampling units for each group - upazila and union for communities engaged under WASH I and WASH II, respectively. For comparison's sake, we used a similar approach for sampling WASH I- and WASH II-engaged communities for our verification exercise. As such, the PSUs for our verification survey were the upazila and union (i.e., smallest rural administrative and local government units) for communities engaged under WASH I and WASH II, respectively. VWCs are the lowest organizational level within BRAC's WASH II programme, and the size of each VWC is relatively similar (there are typically 100-300 households per VWC catchment area). In subsequent stages of sampling (i.e., stage three for WASH I-engaged communities, and stage two for WASH II-engaged communities). See section *1.2.6 Detailing the sampling methodology* for further details regarding the sampling methodology used to select sampling units at each sampling stage.

1.2.3 Identifying ultimate sampling unit

The ultimate sampling unit for the initial verification exercise was the household; specifically, any household residing in an upazila targeted by BRAC's WASH II programme. See the section below entitled *Performing the sample size calculation* for further details regarding selection of households for inclusion in the verification sample.

1.2.4 Ascertaining the sampling frame

It is important to clearly communicate with the implementer from the outset to manage expectations regarding the requirements of the verification sampling frame. The implementation summarizes characteristics of an appropriate sampling frame. For the initial verification, we originally planned to use a list of enumerated households from which BRAC collected QIS data to serve as the QIS sub-sample sampling frame. However, there were concerns regarding the release of personally identifying information. Therefore we were not able to select households into our sample before moving to the field, as planned. Upon arriving in a selected VWC, however, we obtained the VWC register from whomever held the register. The VWC register housed an enumerated list of all households within the VWC catchment area, with household name and wealth category indicated for each household. These registers served as the VWC-specific sampling frames from which households were selected for inclusion in our various sample sub-groups (e.g., QIS WASH I, QIS WASH II, non-QIS WASH I, non-QIS WASH II) via systematic random sampling. See the section below entitled *Performing the sample size calculation* for further details regarding the method used to conduct systematic random sampling within wealth categories at the VWC level.

1.2.5 Performing the sample size calculation

In a recent progress report dated January 2014, BRAC indicated that 75% of poor and hard-core poor households and 80% of non-poor households sampled via its QIS had a latrine that scored at or above benchmark for coverage with a sanitary and hygienic household latrine (BRAC defined the benchmark as latrines which have a ring slab and functioning water seal). The report also indicated that 81% of hard-core poor households with a latrine (hygienic or unhygienic) reported regular use of the latrine, with 77% of households reporting use by its members. Analysis of BRAC's raw QIS data along with the aforementioned reported latrine coverage and utilization proportions formed the basis of our sample size calculation.

In order to determine the total number of clusters required to test our two-sided verification hypotheses, we conducted a power analysis using Monte Carlo simulation by repeatedly simulating the outcomes of interest using model parameters estimated from the BRAC QIS data in SAS. We assumed $\alpha = 0.05$ to assess the power generated by different numbers of clusters for detecting differences in effect between:

1. BRAC's QIS sample and our sample for the three sanitation outcomes of interest (coverage, "use, by" and "use, when") – detectable difference set at +/- 10%;
2. BRAC's QIS sample and our sample for sanitation outcomes of interest, by wealth category; and
3. Wealth categories for sanitation outcomes of interest within our own survey sample.

Rationale for chosen sample size

During its QIS survey, BRAC selected 27 households per VWC cluster, in which a block sample of 9 households per wealth category were sampled (i.e., 9 hard-core poor households, 9 poor households, and 9 non-poor households). In order to determine which households fell within which wealth categories, BRAC used VWC data that were collected during participatory rural appraisals (PRAs) of each community. In the early phases of WASH I / WASH II programs, VWCs conducted the PRAs of their communities, with assistance from BRAC field staff, during which they mapped and classified every household in their community according to BRAC's definition of hard-core poor, poor, and non-poor. The VWCs maintain these maps and documentation of household wealth category classifications in their permanent Committee records. For communities engaged under WASH I, these wealth category assessments took place in 2007, and while new households are added to the community map on an annual basis, the wealth categories of the existing households on the maps are neither routinely nor systematically re-assessed or updated in Committee records. For communities in the expansion upazilas engaged under WASH II, these wealth category assessments took place in 2011. While these wealth category assessments are 3-7 years old, they represent the household wealth category breakdowns BRAC is using for programmatic and analytical purposes.

For the purposes of the verification exercise, we originally assumed 10% of the VWCs contained less than 9 households per wealth category (a stated limitation of BRAC's QIS sampling methodology). We also assumed a 5% attrition rate per cluster for households leaving/moving from their previous QIS locations or other loss to follow-up within the QIS sample. Rounded to the next whole number divisible by 3 (to account for the three wealth categories), this resulted in a desired sample of 21 households per VWC cluster. In order to assess the classification of household wealth and to determine whether the various financing mechanisms are reaching their intended household wealth targets, we initially planned to survey a block sample of 7 households per wealth category (i.e., 7 hard-core poor households, 7 poor households, and 7 non-poor households for a total of 21 households per VWC cluster).[1] During the analysis

[1] Each Village WASH Committee maintains a list of every household in their catchment area, and via a participatory rural appraisal of their communities, has classified all of the households as ultra-poor, poor, or non-poor. While we used the VWC's

phase, we adjusted for block sampling within wealth category strata, by applying a weight factor to each household observation that was equivalent to the inverse of the probability of selection for each household in each wealth category in each VWC.

Subsequent to receiving the raw data from BRAC's QIS survey, we determined that in 61% (182 of 299) of VWC clusters, less than 27 households were surveyed due to fewer than 9 households within the cluster falling into various wealth categories. However, only 9% (26 of 299) of VWC clusters surveyed by BRAC contained less than our targeted 21 households. **Given it was likely that we would sample VWCs with fewer than seven households per wealth category, we inflated our take size by one household per wealth category such that we selected eight households per wealth category via systematic random sampling**.

Households were selected at the VWC level via stratified random sampling. In other words, households were enumerated by wealth category, the sampling interval [k] was calculated by dividing the total number of households in the wealth category by eight, a random number between one and k was generated to identify the first household for selection into the sample, and we subsequently applied the sampling interval from that point forward to select the remainder of the sample for each wealth category. In order to allow for random replacement of households in the field,[2] the team generated a list of 15 randomly selected households per wealth category. During survey administration, enumerators targeted the first eight households per wealth category in each VWC. If one of the first eight households on the list was vacant, had no eligible survey respondent present to represent the household, or refused to participate in the survey, the field supervisor advised the enumerator to replace that household with the ninth household on the list, and so on, until a total of eight households per wealth category were surveyed in each selected VWC. Field supervisors coordinated household replacement. When less than eight households in any given wealth category resided within the VWC, the team surveyed all available and consenting households in that wealth category.

During the calculation of our overall sample size (i.e., during the assessment of the number of clusters that would be needed to generate acceptable power [~80%+] to detect differences in sanitation outcomes at these levels), we decided to base our final sample size on the latrine coverage indicator, as it generated a larger required sample size (i.e., required more clusters to be sampled) than the latrine "use, by" and latrine "use, when" indicators. Going for the largest required sample size would allow us to be adequately powered for all indicators of interest. Our power analysis indicated that 52 clusters (calculations were made with the original take size of 21 households per cluster, for a total sample size of 1,092) would power us to detect a 10% (bi-directional) difference between BRAC's QIS sample and our survey sample for the outcome indicators of interest. This sample size calculation also accounts for the clustered nature of the data.

The raw QIS survey data indicated only a 3% difference in effect for latrine coverage and latrine "use, when" indicators between poor and hard-core poor households. As a result, it would take an extremely large number of clusters to detect differences in effect between these two wealth categories for these indicators. Therefore, we had to make some concessions with the sample size. Due to the fact that BRAC

---

wealth status classifications to select our study sample (as did BRAC to obtain its QIS sample), we assessed the appropriateness of these classifications by collecting information on all criteria BRAC uses to define its wealth categories and comparing the results of our survey to households' wealth classification.

[2] Enumerators were trained to visit the first eight selected households in each wealth category three times prior to replacing that household with the next selected household on the list in the event that a household remains vacant. Households that refuse to participate in the study will be replaced by the next randomly selected household on the list.

reported latrine coverage on two levels (non-poor versus combined poor and hard-core poor) in its progress report, we powered our study to detect differences in effect for coverage between non-poor versus combined poor and hard-core poor between our sample and BRAC's QIS sample as well as within our own verification sample. Due to possible wealth category misclassification (given BRAC household wealth category assessments are 3-7 years old), as well as monetary and time restrictions, we decided to power our study to detect differences in latrine "use, when" for non-poor versus poor and hard-core poor (as an aggregate) between our sample and BRAC's QIS sample as well as within our own validation sample. Our study is also powered to detect differences in latrine "use, by" for non-poor versus poor and hard-core poor (as an aggregate) as well as poor versus hard-core poor between our sample and BRAC's QIS sample as well as within our own verification sample. While we will be able to report descriptive statistics for our QIS sub-group and our non-QIS sub-group, our study is not powered to detected differences in effect between these two sub-groups. Our study is also not powered to detect differences in latrine coverage and "use, when" for poor versus hard-core poor between our sample and BRAC's QIS sample or within our own verification sample.

In keeping with BRAC's reported sampling strategy (selecting half of the sample from WASH I-engaged clusters, and half of the sample from WASH II-engaged clusters), we selected the 52 required clusters equally from QIS and non-QIS groups, and equally from WASH I and WASH II-engaged upazilas/unions within the QIS and non-QIS groups. In other words, we selected 13 clusters (312 households) from WASH I-engaged communities and 13 clusters (312 households) from WASH II-engaged communities from the QIS sampling frame to comprise the QIS group sub-sample, and 13 clusters (312 households) from WASH I-engaged communities and 13 clusters (312 households) from WASH II-engaged communities from the non-QIS sampling frame to comprise the non-QIS group sub-sample.

1.2.6 Detailing the sampling methodology

During the initial verification exercise, our team employed a multi-stage sampling strategy to select the 21 households per cluster. During the first stage, we selected 26 PSUs from WASH I-engaged communities (i.e., we selected 13 upazilas from the QIS sampling frame to help comprise our QIS sub-group, and 13 upazilas from the non-QIS sampling frame to help comprise our non-QIS sub-group), and we selected 26 PSUs from WASH II-engaged communities (i.e., we selected 13 unions from the QIS sampling frame to help comprise our QIS sub-group, and 13 unions from the non-QIS sampling frame to help comprise our non-QIS sub-group) using probability proportional to size (PPS). During the second stage for WASH I-engaged communities, we selected one union per upazila using PPS. We selected VWCs from each selected WASH I- and WASH II-engaged union via simple random sampling. In the final stage, we obtained the VWC register (VWC-level household sampling frame) from selected VWCs, and then stratified the sampling frame by wealth category as defined by BRAC/VWC PRA data (i.e., ultra-poor, poor, non-poor).

The inclusion and exclusion criteria utilized for the initial verification activity are outlined below.

Inclusion criteria: All households situated within in the 157 WASH II program upazilas were eligible for inclusion into the verification sample. As previously mentioned, households from each wealth category were sampled to determine the accuracy of wealth category classification and receipt of targeted pro-poor/hard-core poor programme interventions. In order for households to be selected for inclusion in our verification sample's QIS group sub-sample, the household must have resided in a VWC that was included in BRAC's QIS survey sample. Researchers ensured household eligibility by selecting VWCs for the QIS group sub-sample from the list of VWCs BRAC sampled during its QIS survey. In order for households to be selected for inclusion in our verification sample's comparison group sub-sample, the household must have resided within one of the 102 upazilas not included in BRAC's QIS survey sample. In order to be eligible for the survey, the household must have had at least one adult who consented to participating in the verification exercise who served as the primary survey respondent. Further, the enumerator must also have deemed the adult individual capable of understanding and providing informed consent.

Exclusion criteria: Any household selected for our verification sample that refused to be surveyed, or either was repeatedly vacant after three attempts within the course of a day or did not have an appropriate member of the household (capable female or male 18 years of age or older) home to serve as the household's respondent was excluded from the verification sample. These households were replaced by the next randomly selected household within that particular wealth category in the VWC.

2. RECRUITMENT & TRAINING

During the initial verification exercise, survey enumerators, field supervisors, and PLUM team members were recruited from icddr,b's pool of trained contractors and staff. The individuals had previous experience working with icddr,b and possibly other organizations/institutions in administering surveys, particularly those focused on water, sanitation, and hygiene. The majority of survey enumerators recruited for this survey had recently completed collection of data for Bangladesh's National Hygiene Survey.

Once recruited, staff from the implementing organization who were responsible for managing the QIS monitoring system provided a two-day training (exclusive of field testing) on the QIS approach. During the two-day QIS training, staff from the implementing organization first explained the general ladder scale of scoring responses, and then went into more detail to explain each QIS survey prompt. All field staff participated in interactive exercises to enhance understanding of both the survey prompts and the QIS scoring process. After the didactic training and interactive activities, trainers answered outstanding questions field staff had regarding the QIS survey, and finally the trainers and field staff discussed examples of situations and what the appropriate scores would be under those circumstances.

Subsequent to the two-day QIS training, the Emory University Research Project Manager and senior icddr,b staff conducted a two-day training (exclusive of field testing) on all remaining survey modules. The Research Project Manager and senior icddr,b staff also re-trained survey enumerators, field supervisors, and PLUM team members on research ethics, interview techniques, and field procedures. The Research Project Manager and senior icddr,b project staff provided enumerators and field supervisors a detailed description of questionnaire content during which the trainers and field staff reviewed the specific purpose of each survey question and discussed all possible answer selections. The training was an open forum, with field staff who had considerable experience not only in survey administration in general, but also WASH surveys in particular. Therefore, trainers accepted feedback from field staff, and adjusted the survey instrument accordingly prior to field testing.

3. DATA COLLECTION PROCEDURES

For the purposes of the initial verification, data collection involved reviewing data from existing data sources, observing programme operations and interventions, administering surveys, and conducting other activities such as latrine spot checks and sensor deployment to collect latrine utilization data. During our verification exercise, latrine coverage was ascertained through household surveys and verified by visual inspections that provided information on latrine completeness, condition, and maintenance. As latrine utilization is especially difficult to confirm, and empirical evidence indicates utilization cannot be assumed from coverage, we employed multiple methods for measuring utilization that have shown promise in our work in Orissa, India. The team thus collected data on latrine use through three measurement techniques:

1. administration of household use schedules (i.e., questionnaires) capturing self-reported latrine use data for all household members;
2. direct observation of latrines during which spot check indicators were assessed to determine use; and
3. instrumented monitoring capturing sensor-recorded use data via passive latrine use monitor (PLUM) deployment to household latrines.

3.1 Administering the survey

A team of enumerators trained by icddr,b senior supervisory staff and the Emory study coordinator administered the survey at randomly selected households. Survey administration took place during 28 June and 8 September 2014, after the receipt of ethics board approvals from Emory University IRB and icddr,b RRC and ERC. Enumerators sought out adult respondents, with preference going first to the primary female caretaker of the youngest child within the household (as she would tend to know the most about the latrine use and defecation practices of most members within her household). If she was not available, enumerators sought out other household members to serve as primary survey respondents using the following preferred hierarchy: eldest available female caretaker, eldest available female household member, eldest available male caretaker, or eldest available male household member. All household members present during survey administration were asked to self-report on their own latrine use habits. If after three attempts at a household, no adult respondent was available to answer the survey, the household was replaced by the next randomly selected household for that wealth category on a previously generated list of randomly selected households. Household questionnaires captured information related to respondents' demographics, socioeconomic status, latrine attributes (e.g., latrine construction, maintenance, and repair), self-reported use, self-reported programme software exposure, the three indicators developed by BRAC and measured via its QIS (i.e., QIS indicators), and self-reported experience with recent illness/symptoms.

QIS indicators

In order to ensure comparability of QIS indicator collection and to limit bias between BRAC's QIS survey administration and our survey administration of the three QIS indicators of interest, we invited BRAC trainers to train icddr,b enumerators on how to administer the QIS survey prompts and how to scale household scenarios via BRAC's QIS 'ladders'. BRAC trainers oriented our enumerators to the QIS administration and scaling approaches, and enumerators practiced employing these approaches during practice sessions. BRAC trainers were also on hand to provide additional guidance during pilot testing.

Enumerators electronically captured survey data and direct observations on password-protected hand held computer tablets. The tablets included programming that allowed for range and consistency checks. During the final stages of the survey, enumerators measured distances between the latrine(s) and various

points of interest (e.g., the homestead, the nearest water source). Field supervisors and study coordinators were on hand to supervise survey administration.

3.2 Performing direct observation via latrine spot checks

During the initial verification exercise, enumerators assessed a series of indicators related to latrine structure, functionality, cleanliness, and general indicators of utilization. Enumerators captured this information from each latrine used by the household, starting with the most frequently used latrine, moving to the second most frequently used latrine, and so on until data were captured for each household latrine. As the enumerators directly observed various aspects of each latrine, they requested that the survey respondent move with them to each latrine to provide detail about the ownership of the latrine (e.g., household ownership, shared ownership) , the users of the latrine (e.g., household members only, shared with others outside of the household), the date the latrine was most recently constructed/repaired, whether the household received any outside assistance for that construction/repair, which entity provided the assistance and to what extent, and what/if any contribution the household made with regard to labor. It was important for the survey respondent to move with the enumerator to each latrine so the respondent was not confused about the specific latrine that was being discussed (in the event that the household had multiple latrines). Having the respondent go with the enumerator to visit each latrine also indicated the typical route the respondents would take to the latrine, which assisted the enumerator in identifying the route to measure to capture the distance between the latrine and the homestead.

After obtaining information regarding latrine construction and repair, the enumerator was prompted via computer tablet to observe and record various aspects of the latrine related to its structure, function, and apparent indicators of use. Further information regarding specific latrine spot check indicators is provided below in Table 3.

3.3 Deploying sensors to detect latrine utilization

PLUMs have never been deployed in Bangladesh, and we were uncertain of the household refusal rate. We anticipated that seeking to obtain PLUM data from slightly over one-half of QIS VWCs would result in PLUM data from at least one-third of QIS VWCS after accounting for data loss (typical with PLUMs, particularly during monsoon season) and household level refusal of PLUM installation. During the one-week data collection period, the PLUM collected household level latrine use measures via its remote sensing capabilities.

All consenting households within a sub-set of randomly selected VWCs had a PLUM placed in all functional household latrines for one week, such that approximately one-half of randomly selected VWCs in the QIS group sub-sample (i.e., 14 VWCs [7 WASH I and 7 WASH II VWCs], for a total of 336 households) were requested to provide PLUM data. In order to randomly select QIS VWC clusters to provide PLUM data, we calculated a QIS VWC sampling interval, used a random number generator to locate our starting point, and applied the sampling interval to identify the first randomly selected PLUM VWC cluster. We continued with this strategy until we obtained a list of 14 randomly selected PLUM clusters.

Senior research staff from Emory and icddr,b trained a specialized team of field staff in PLUM installation and removal. To the greatest extent possible, PLUMs were installed in the early morning hours, and uninstalled in the evening hours. Enumerators fully exited the latrine just prior to indicating the PLUM installation time so as to not trigger the PLUM sensors to indicate a latrine event occurring at that moment. Data are stored in internal temporary memory stores within the PLUMs for 24-hours prior to

being transferred to an internal microSD card at mid-night each day. Given low mobile network coverage in many areas we administered our survey, we did not attempt to use an operational mode to upload data to PSU servers in real-time. After approximately one week, the PLUM team went back to each household latrine to retrieve the PLUMs. Just prior to entering the latrine to retrieve the PLUM, the PLUM team member recorded the date and time of PLUM retrieval.

3.4 Drafting a data dictionary

During the initial verification exercise, the verification team created a data dictionary/SOP that organized variables by module, in order of their appearance in their respective modules, and provided details on the following aspects of each variable:

- Survey question number
- Variable name in the dataset (this remained blank until the team extracted data from the tablets)
- Description of the variable
- Value of answer selection (e.g., numeric, alpha numeric, categorical, including all possible answer choices for categorical responses)
- Description of value (i.e., how the answer should be obtained and/or confirmed – e.g., by asking, by enumerator observation), including the codes for all possible answer choices
- Skip notes (i.e., notes that indicate possible skips in the order/presentation of survey questions) are important for documenting skip patterns within the survey instrument, and ensuring the patterns are complete and make sense in the context of the survey instrument (this information will assist with the missing data analyses)
- Objectives of the variable/survey prompt
- Additional remarks