

Report on a Study to Independently Assess Latrine Coverage and Use under BRAC's WASH II Project in Bangladesh

Developed by

Department of Environmental Health
Rollins School of Public Health
Emory University

Department of Mechanical and Materials Engineering
Portland State University

Water, Sanitation, and Hygiene Research Group
Centre for Communicable Diseases
International Centre for Diarrhoeal Disease Research, Bangladesh

FINAL REPORT

April 2015



EMORY
UNIVERSITY

**Center for
Global Safe Water**

For more information, contact:

Maryann Delea
Department of Environmental Health
Rollins School of Public Health
Emory University
1518 Clifton Road NE
Atlanta, GA 30322

Email: mdelea@emory.edu

ACKNOWLEDGEMENTS

This study would not have been possible without the work of many dedicated colleagues. Special thanks to Bimal Kumar Das, Shiuli Das, and the team of field supervisors and survey enumerators who contributed to the initial verification exercise.

Contributors of this report include:

Emory University

Rollins School of Public Health

Departments of Environmental Health and Biostatistics & Bioinformatics

Maryann Delea, MPH

Matthew Freeman, PhD, MPH

Howard Chang, PhD

Sabah Ghulamali, MPHc

Michael Austin, MPHc

Thomas Clasen, PhD, JD

International Centre for Diarrhoeal Disease Research, Bangladesh

Centre for Communicable Diseases

Water, Sanitation, and Hygiene Research Group

Amal Krishna Halder, PhD

Abul Kasham Shoab

Nuhu Amin, MPH, BDS

Probir K Ghosh

Mamun Ar-Rashid, MPH

Supta Sarker

Leanne Unicomb, PhD

Portland State University

Department of Mechanical and Materials Engineering

Zak White, BSCE

Evan Thomas, PhD, PE, MPH

Support for this study was provided by The Bill & Melinda Gates Foundation, Global Development Grant Number OPP1102989.

CONTENTS

LIST OF ACRONYMS.....	6
EXECUTIVE SUMMARY	7
1. INTRODUCTION.....	13
1.1 BRAC’s Qualitative Information System.....	13
1.2 Verification objectives	14
1.3 Verification questions	14
2. METHODS.....	15
2.1 Verification design	15
2.1.1 Target verification population	15
2.1.2 Sampling frame.....	15
2.1.3 Primary and secondary sampling units.....	15
2.1.4 Ultimate sampling unit	16
2.1.5 Sample size calculation	16
2.1.6 Sample selection.....	16
2.2 Outcome measurement.....	19
2.2.1 Latrine coverage	19
2.2.2 Latrine utilization	21
2.2.3 Latrine utilization measurement method cost assessment	23
3. RESULTS	24
3.1 VERIFICATION QUESTION 1.....	25
3.1.1 Latrine coverage	25
3.1.2 Latrine utilization	30
3.2 VERIFICATION QUESTION 2.....	44
3.2.1 Intervention receipt by wealth category: self-reported receipt of outside assistance	44
3.3 VERIFICATION QUESTION 3.....	45
3.4 VERIFICATION QUESTION 4.....	45
3.4.1 Comparisons of latrine utilization measurement methods.....	45
3.5 Cost assessment for various latrine utilization measurement methodologies	47
4. DISCUSSION.....	48
4.1 Comparability of results.....	48
4.2 Wealth category designations	49
4.3 Reported latrine use	49
4.4 Directly observed indicators of use	50
4.5 Instrument-recorded use.....	51
4.5.1 PLUM data limitations	51
4.6 Latrine utilization measurement method cost considerations.....	51
5. RECOMMENDATIONS	52
REFERENCES.....	54

APPENDICES

- APPENDIX I Additional verification details
- APPENDIX II Verification data tables
- APPENDIX III Analysis of BRAC's WASH II monitoring system

LIST OF ACRONYMS

ARI	Acute respiratory infection
CI	Confidence interval
DGIS	Dutch Ministry of Foreign Affairs
EKN	Embassy of the Kingdom of The Netherlands
ERB/ERC	Ethics review board/Ethics review committee
GEE	Generalized estimating equations
GPS	Global positioning system
icddr,b	International Centre for Diarrhoeal Disease Research, Bangladesh
IQR	Interquartile range
IRB	Institutional Review Board
IRC, Netherlands	International Water and Sanitation Resource Centre
M&E	Monitoring and evaluation
MIS	Management Information System
OD	Open defecation
PI	Principal investigator
PLUM	Passive latrine use monitor
PPS	Probability (of selection) proportional to size
PRA s	Participatory rural appraisals
PSUs	Primary sampling units
PSU	Portland State University
QIS	Qualitative Information System
RD	Risk difference
RFP/A	Request for proposals / Request for applications
RR	Risk ratio
RRC	Research review committee
SOP	Standard operating procedure
SRS	Simple random sample
SSUs	Secondary sampling units
USUs	Ultimate sampling units
VWC	Village Water, Sanitation, and Hygiene (WASH) Committee
WASH	Water, sanitation, and hygiene

EXECUTIVE SUMMARY

Background. The Bill & Melinda Gates Foundation (the Foundation) commissioned a study intended to verify programmatic results reported by BRAC, a Foundation grantee implementing a large-scale water, sanitation, and hygiene (WASH) project in Bangladesh. The Foundation requested that our verification team design and conduct a study to verify sanitation outcomes in a cost-effective manner that did not duplicate the implementer's existing monitoring and evaluation activities. The purpose of the study was to verify the accuracy of relevant data by spot checking a random sample of households and assessing latrine coverage and latrine utilization outcomes. The Foundation was interested in verifying implementer-reported results from year one, but also requested that we obtain the implementer's most up-to-date monitoring data. As such, the Foundation requested we review BRAC's monitoring data, as well as the methods used to collect information on relevant indicators, and select a random sample of households from which monitoring data were already collected. The Foundation requested that we verify latrine utilization data through the following methods:

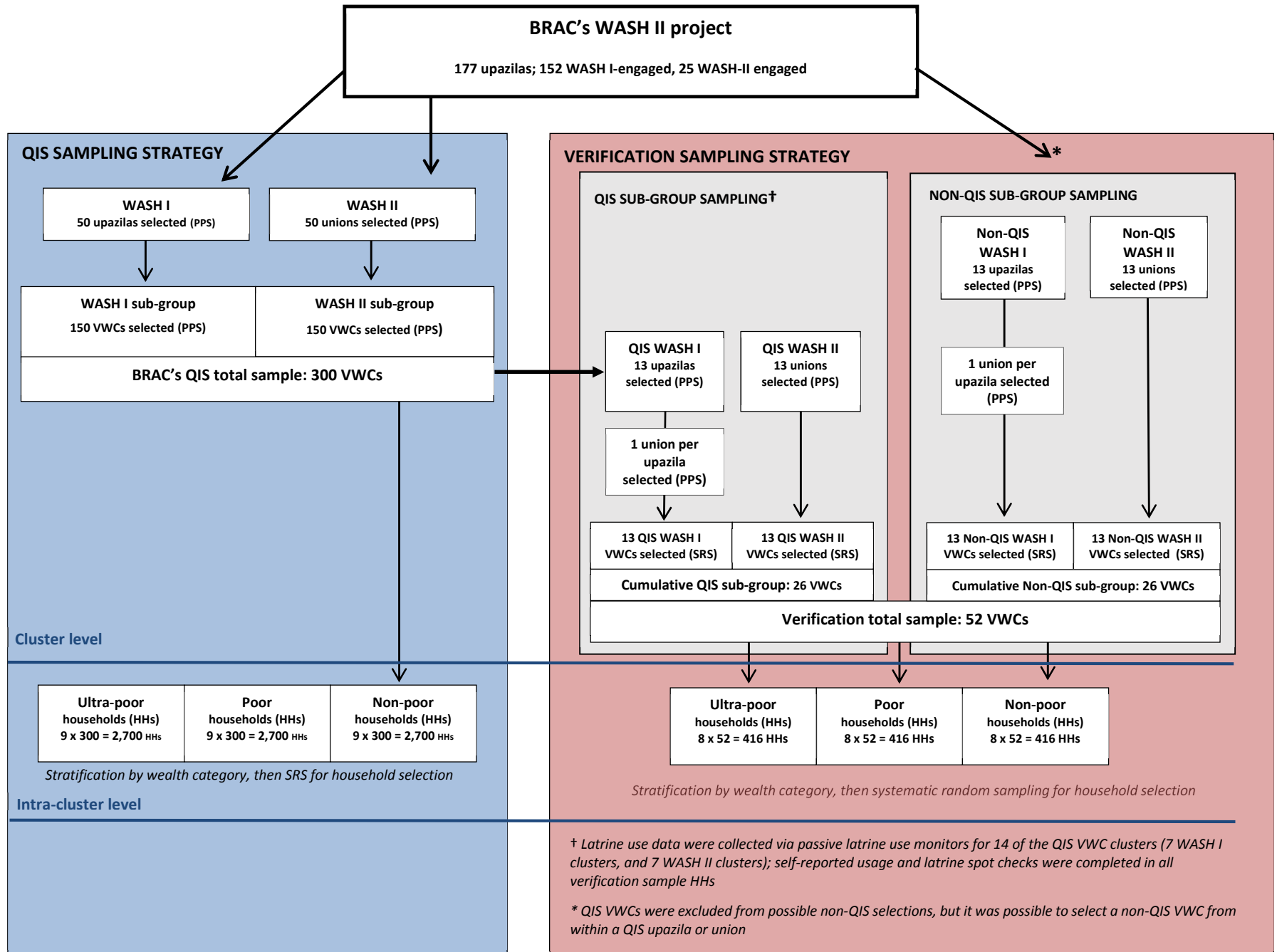
1. survey administration (capturing self-reported use data);
2. direct observation; and
3. sensor deployment (capturing sensor-recorded use data) (Bill & Melinda Gates Foundation 2013).

Sampling strategy. In order to ascertain whether BRAC's QIS sample, selected from 299 Village WASH Committees (VWCs) within 178 unions (165 WASH I, 13 WASH II) and 72 upazilas (sub-districts - 69 WASH I, three WASH II), is representative of its larger project population, we surveyed a comparison group selected from VWCs within the 105 upazilas not included in BRAC's QIS sample. We selected an approximately similar number of households from QIS (VWCs within BRAC's QIS sample) and non-QIS VWCs (comparison VWCs from the larger WASH II project population, but outside of BRAC's QIS sample). The enumerated list of VWCs monitored by BRAC through its QIS served as the sampling frame for our QIS group sub-sample. The enumerated list of all VWCs not included in BRAC's QIS sample served as the sampling frame for the non-QIS comparison group sub-sample.

Verification staff employed a multi-stage sampling strategy to select the 24 households per selected VWC cluster. During the first stage, we selected 26 primary sampling units (PSUs) from WASH I-engaged communities (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 (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 group), using probability proportional to size (PPS). During the second stage in WASH I communities, we selected one union per upazila, via PPS. During the next sampling stage for WASH I and WASH II communities, we selected one VWC from each union via simple random sampling (SRS). In the final stage, we obtained the VWC register (household sampling frame) from selected VWCs, and stratified the sampling frame by wealth category as defined by BRAC/VWC (i.e., ultra-poor, poor, non-poor). Finally, 8 households were selected per wealth category via systematic random sampling.

Figure 1 depicts the sampling strategies employed for the BRAC QIS survey (per BRAC's QIS documentation, though, there was a deviation from the documented strategy) and our verification study.

Figure 1. Sampling strategy – QIS and verification samples



Sources of data. For the purposes of this verification, data collection involved reviewing data from existing WASH II monitoring data sources, observing program operations and interventions, administering surveys, and conducting other activities such as latrine spot checks and sensor deployment to collect latrine utilization data. Latrine coverage was ascertained through household surveys, and was further verified by visual inspections that provided information on latrine completeness, condition, use, and maintenance. As latrine utilization is especially difficult to confirm, and empirical evidence indicates utilization cannot be assumed from coverage (Banda et al. 2007; Montgomery et al. 2010), we employed multiple methods for measuring utilization that have shown promise in our work in Orissa, India (Clasen et al. 2012a). The team thus collected data on latrine use through three measurement techniques:

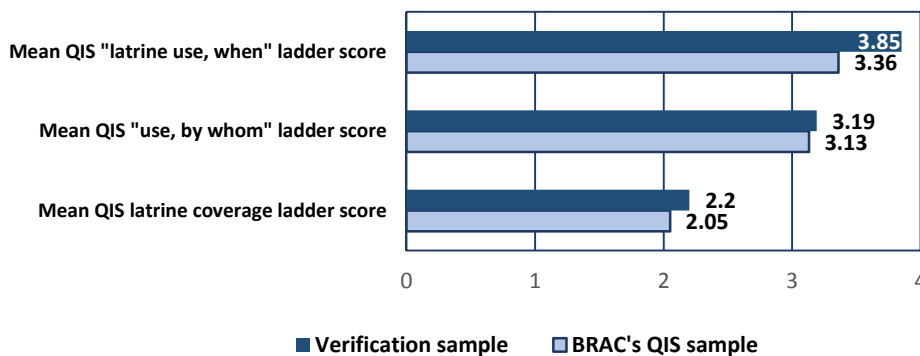
1. administration of household use schedules (i.e., questionnaires) capturing self-reported latrine use data for all household members;
2. direct latrine observation during which spot check indicators were assessed to determine use; and
3. instrumented monitoring, capturing electronically-recorded use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

Our verification of BRAC’s data included a comparative analysis of latrine coverage and utilization proportions reported in BRAC’s narrative progress report as well as a direct comparison of QIS ladder scale distribution data between BRAC’s QIS monitoring dataset and our verification dataset for all three of BRAC’s QIS latrine coverage and use indicators. We compared BRAC’s QIS data to sanitation outcome measures obtained via our administration of BRAC’ QIS ladder scaling system as well as survey and observation data.

In order to ensure comparability of data for our comparative analyses, we incorporated a separate module in our household questionnaire in which we administered the three relevant QIS survey prompts. To minimize survey administration bias, we invited trainers from BRAC to train icddr,b enumerators on how to administer the QIS survey prompts and how to scale household scenarios via BRAC’s QIS “ladders”. In order to further limit administration bias, particularly on the QIS questionnaire module (i.e., enumerators adopting a more systematic strategy for inquiring about latrine utilization for all members of the household), we designated specific enumerators to administer the QIS module only. We selected our female enumerators to administer the QIS module, as it is culturally more appropriate for women to speak more openly with other women than with men.

Survey results. With some exceptions, our verification survey results compare reasonably well with BRAC’s reported results. Figure 2 provides population-averaged summary estimates of mean QIS ladder scores for QIS sanitation outcome indicators in surveyed areas.

Figure 2. Population-averaged mean QIS ladder score estimates for BRAC QIS and verification samples



In general, the QIS ladder scores for latrine coverage and utilization generated from our verification sample were higher than those generated from BRAC's QIS sample. These increases in scores may be due to the amount of time that passed between BRAC's first round of QIS data collection and our verification data collection. During the year between data collection time points, BRAC continued implementing its WASH II project, during which progress was made to improve sanitation outcomes, and household-level attributes (e.g., income, education, and migration) may have changed. In addition to the lapse of time between data collection points, seasonality may also have contributed to the differences in latrine coverage and utilization measures. There was less rain during the 2014 monsoon season compared to prior years, and the overall security situation was stable during our data collection period. Both the presence of rain and insecurity often results in decreased use of latrines, particularly during the nighttime. These temporal changes may account for some of the increase in reported latrine utilization. Although decreases in metrics were rare, they were observed. These decreases may be explained by the destruction of hardware resulting from flooding, natural disasters, and poor household maintenance.

Latrine coverage. In assessing reported latrine coverage benchmarks, as indicated by BRAC in their narrative progress report, no significant differences were observed in reported latrine coverage at or above benchmark amongst non-poor households. However, a significantly higher proportion of verification QIS sub-group households were at or above benchmark compared to BRAC's QIS sample (67.7 vs. 57.0%, respectively; Risk Difference [RD]=9.7, 95% CI 0.8, 18.5).

Reported latrine use. In assessing the reported "latrine use, by whom" benchmark, as indicated by BRAC in their narrative progress report, a significantly higher proportion of ultra-poor verification households scored at or above benchmark compared to BRAC's ultra-poor QIS households (87.6 vs. 75.4%, respectively; RD=12.1, [7.2, 16.9]). Similarly, in assessing the reported "latrine use, when" benchmark amongst ultra-poor households, as indicated by BRAC in their narrative progress report, a significantly higher proportion of ultra-poor verification households scored at or above benchmark compared to BRAC's ultra-poor QIS households (95.6 vs. 80.7%, RD=14.6 [10.9, 18.3]).

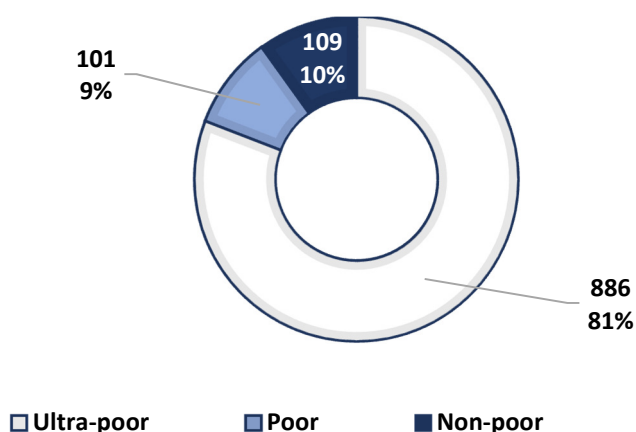
Other measures of use. Some evidence of open defecation (i.e., human feces) was observed in 55.8% (29/52) of surveyed clusters. Several latrine spot check indicators are associated with either of BRAC's two QIS latrine use indicators (i.e., "latrine use, by whom" and "latrine use, when"). Three latrine spot check indicators – visible discoloration of latrine pan or slab, available water for flushing or anal cleansing, and available water near latrine for handwashing – are associated with both QIS use indicators. After accounting for data loss due to the variety of reasons, PLUM data were captured on 217 primary (i.e., most frequently used household latrine, per household respondent) latrines, and 14 secondary (i.e., second most frequently used household latrine) latrines from 220 households within the QIS sub-group. After adjusting for survey design, households that had a PLUM installed secondary to survey administration were found to self-report a four-day average of 32.8 events (95% CI 28.6, 37.0) vs. a four-day average of 21.7 events (95% CI 18.1, 25.4) recorded with the PLUMs. This suggests over-reporting of self-reported latrine utilization.

Intervention receipt by wealth category. Our results indicate a strong association between wealth status and reported receipt of latrine construction assistance from BRAC during the WASH II project period ($p < 0.001$), as population-averaged statistics indicate that 48.9% (886/1811) of ultra-poor, 6.2% (101/1625) of poor, and 6.2% (109/1765) of non-poor households with an improved or shared but otherwise improved latrine reported receiving latrine construction assistance from BRAC since 2011. These figures indicate that approximately half of all ultra-poor households in surveyed areas reported receiving latrine construction assistance from BRAC during the WASH II program period. This is consistent

with what BRAC reports as its WASH II program target for direct hardware intervention support (i.e., latrine construction assistance).

Figure 3 presents population-averaged estimates for self-reported latrine construction assistance from BRAC for the construction of an improved or improved but otherwise shared latrine during the project period for areas surveyed during the verification. These figures indicate that the majority (81%) of the households in surveyed areas that have an improved or shared but otherwise improved latrine (i.e., the type of latrines BRAC promotes under the WASH II program) and reported receiving latrine construction support from BRAC are ultra-poor households. While BRAC aims to target only ultra-poor households with direct latrine construction support, a small proportion of non-poor and poor households with an improved or improved but otherwise shared latrine reported receiving latrine construction support from BRAC (10% and 9% for non-poor and poor households, respectively). These reports may indicate a misdirection of program interventions, a misclassification of household wealth status at the VWC level, or respondent recall bias.

Figure 3. Population-averaged estimates for receipt of BRAC latrine construction support amongst households with an improved or shared but otherwise improved latrine



Cost assessment for various latrine utilization measurement methodologies. Our cost assessment revealed that instrument-recorded latrine utilization was the most expensive latrine utilization measurement method. These instrument-recorded data incur approximately five times the administration costs and nearly two and a half times the equipment rental costs of those incurred to obtain latrine spot check indicator data, and just over one and a half times of those incurred to obtain self-reported use data.

Recommendations. While our verification results compare reasonably well with BRAC's reported results, there is room for improvement of indicators and measurement methodologies. The following is a summary of recommended modifications. See section 5. *Recommendations* for further details.

1. Strengthen and harmonize monitoring systems:

- **Centralize data availability:** While it is important to manage and utilize project data at field and regional office levels, it is also important to ensure data flows to, and are maintained at the central office level. We observed that high level project data (e.g., a list of all VWCs in which the project is operating, a codebook for the QIS monitoring system dataset) were not readily available at the central office. This highlighted issues with data flow and management within the project monitoring system.

- Maintain a centralized list of all VWCs in which the WASH II project is operating. An enumerated list of all areas in which a project is operating is important not only from a program management perspective, but also a project monitoring and evaluation perspective.
 - Maintain a complete codebook of monitoring system data. The availability of a codebook will not only ensure proper interpretation of data amongst central office and M&E sub-contracting staff, but also allow future verification entities to properly analyze and interpret verification data.
 - **Household identifiers:** The current monitoring system uses a different MIS and QIS household identifier for the same household. Streamlining simple aspects of the two disparate monitoring systems will serve to improve monitoring and reporting.
 - **Electronic data capture:** While the MIS system electronically captures information regarding the number of latrine constructed and repaired, latrine utilization data captured on household follow-up (after receipt of hardware subsidy/loan) is not electronically entered into the database, and hardcopies of monitoring reports sent from regional offices to the central office are destroyed after three months. Consider entering latrine utilization data captured in MIS hard copy reports into the electronic MIS database.
 - **Cross-check data prior to circulation:** One of the Excel spreadsheets containing data on intervention implementation was inaccurate – it under-reported subsidy and loan distribution because the cell formulas were not correct. In order to prevent future under- or over-reporting, databases should be cross-check for accuracy.
- 2. Incorporate monitoring indicators to align with the Foundation’s specified outcome indicators:**
- **Intervention receipt and sustained use indicators:** Consider including simple survey prompts within the QIS instrument that inquire households about when household latrines were constructed and/or last repaired (to allow program staff to determine the proportion of adults using the latrine 6 months after installation/repair), whether, when, and by whom a subsidy or loan was received by the household for latrine construction/repair. These four prompts should add no more than 2 minutes to QIS survey administration time, but would allow program administrators to cross-check MIS data on intervention receipt, and make statements about the sustainability of latrine utilization subsequent to installation/repair.
 - **Incorporate objective indicators of latrine utilization into QIS latrine use ladder scores:** As evidenced by the comparative analysis of self-reported and instrument-recorded data, self-reported latrine utilization data are often inflated. While it may not be practical to install sensors in latrines throughout the life of the WASH project, the inclusion of other objective indicators of latrine use may improve the veracity of self-reported latrine use.
 - Consider incorporating some of the latrine use indicators most strongly associated with use as criteria within QIS ladder score definitions.
- 3. Systematically re-assess wealth category on a routine basis:** VWC wealth category classifications are 2-7 years old, and though new households have been added to VWC registers and community maps, a systematic re-assessment of household wealth category designation has not taken place. As WASH II project intervention implementation is informed by household wealth status, it is important that the household wealth designation is accurate and up-to-date.

1. INTRODUCTION

In 2011, the Bill & Melinda Gates Foundation (the Foundation) awarded a grant to BRAC to expand the implementation of a sustainable sanitation delivery model in rural Bangladesh, with a specific focus on the very poor. This grant from the Foundation, and another from the Embassy of the Kingdom of The Netherlands (EKN/DGIS) support a large-scale water supply, sanitation and hygiene (WASH) promotion project known as “WASH II”, a project that operates in 177 of Bangladesh’s 492 upazilas (sub-districts). During August 2011, BRAC, the International Water and Sanitation Resource Centre (IRC, Netherlands – a project collaborator), DGIS, and the Foundation agreed to treat WASH II as a single project (Sijbesma et al. 2012).

Measurement of outcomes is central to the Foundation’s sanitation strategy. In its partnership with BRAC, the Foundation’s disbursement of funds is tied to the achievement of scale and pro-poor targets, rendering independent verification critical. In early 2014, the Foundation commissioned Emory University to conduct an independent assessment of sanitation outcomes reportedly achieved by BRAC’s WASH II project in Bangladesh. Though the Foundation’s primary interest was in assessing the accuracy of latrine coverage¹ and use² reported by BRAC, it was also interested in assessing whether the project is reaching its pro-poor (including ultra-poor and poor households) intervention targets. The Foundation also sought to compare the results and respective methodological costs of surveys of self-reported household³ latrine use with those based on spot checks and instrumented monitoring. As a result, Emory University subsequently engaged collaborators from the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) and Portland State University (PSU) to support field data collection activities and passive latrine use monitor (PLUM) technical assistance, respectively.

1.1 BRAC’s Qualitative Information System

BRAC’s internally collected monitoring data served as the basis of comparison for this verification study. A full analysis of BRAC’s WASH II project monitoring system and comprehensive details related to BRAC’s Qualitative Information System (QIS) can be found in Appendix III. The description provided herein is a brief synopsis of BRAC’s QIS in relation to the objectives of this particular verification exercise.

BRAC routinely monitors all households receiving direct intervention support for latrine hardware via their WASH II grants/subsidies, loans, and latrine repair services. Grants/subsidies for latrine construction are targeted toward ultra-poor households, and micro-finance/revolving loans are targeted toward poor households. Latrine repair services are available for both poor and ultra-poor households. In addition to

¹ In order to ensure comparability between survey indicators, the verification explored two levels of latrine coverage: 1) the proportion of households with access to a latrine, as measured by both the QIS latrine coverage indicator and direct observation; and more specifically, 2) the proportion of households with a latrine constructed or repaired as a result of direct and/or indirect WASH II project interventions (indicator to be verified, per the Foundation RFP).

² In order to ensure comparability between survey indicators, the verification explored latrine utilization through four different methodologies: 1. BRAC’s QIS ladder scaling, in which household latrine use is scored according to the individuals using the latrine (i.e., “latrine use, by whom” [HH04 indicator, per QIS]), and when these individuals use the latrine (i.e., “latrine use, when” [HH05 indicator, per QIS]); 2. self-reported latrine use data captured via a household use schedule of all latrine users utilizing the household latrines; 3. instrument-recorded latrine use utilizing a passive latrine use monitor (PLUM) that captured defecation events over a four-day analytical period; and 4. latrine spot check indicators captured via direct observation of each household latrine. The verification also examined the extent of reported use for each member of the household and each non-household member who uses the household latrine on a regular basis.

³ For the purposes of the verification, a household is defined as a person/group of related/unrelated persons who usually live together in the same dwelling(s), who have common cooking/eating arrangements, and who acknowledge one adult member as head of household.

their routine monitoring system that tracks primarily process indicators (e.g., number of grants and loans distributed and received, number of latrines constructed, number of latrines repaired), BRAC and the International Water and Sanitation Resource Centre (IRC, Netherlands) conduct a survey on a sample of households once-per-year via its QIS. The sample of households monitored through the QIS approach is comprised of households receiving direct and indirect support from WASH II. Indirect intervention support consists of the “software” of the project, and centers on awareness raising and mobilization of households to improve their sanitation situation in the absence of the receipt of direct financial assistance for doing so.

Through the QIS system, BRAC is able to quantify qualitative process and outcome indicators via progressive scales or ‘ladders’ for seven WASH indicators (two of which are related to safe water, four to sanitation, and one to hygiene). Each step in the ladder has a short description, called a mini-scenario, which describes the situation for a particular indicator score. The Foundation asked Emory to develop and conduct a study that collects data to verify the sanitation outcomes obtained via BRAC’s QIS monitoring system.

1.2 Verification objectives

The objectives of this verification were as follows:

1. Assess the accuracy of BRAC’s reported latrine coverage and usage rates by comparing BRAC’s QIS data to a cross-sectional household survey and instrument (PLUM)-recorded latrine utilization data;
2. Determine which households are receiving grants/subsidies for latrine construction and latrine repair services in order to ascertain whether various financing mechanisms are reaching their intended household wealth targets, per their socio-economic status classifications;
3. Determine whether latrine coverage and use varies based on implementation strategy;
4. Identify the most cost-effective techniques for assessing latrine use by presenting costing information for the following data collection methods:
 - a) survey data capturing self-reported use
 - b) latrine spot checks (i.e., visual inspections of the latrine and surrounding premises)
 - c) electronically logged data collected via passive latrine use monitors (PLUMs); and
5. Summarize and identify potential strengths and weaknesses of BRAC’s monitoring systems for latrine construction/repair and use, including its sampling strategy.

1.3 Verification questions

The following research questions were drafted in order to attend to the above-mentioned verification objectives.

1. Are BRAC’s monitoring data accurately reflective of latrine coverage and use amongst BRAC’s WASH II project participants?
2. Is BRAC achieving its poverty targeting goals (i.e., are ultra-poor households receiving grants/subsidies for latrine construction/repair, and poor and ultra-poor households receiving micro-finance/revolving loans for latrine construction/repair)?
3. Are latrine coverage and utilization similar among project participants receiving support via grants/subsidies compared to those receiving micro-finance loans?
4. How do various measures of latrine utilization (i.e., QIS ladder scales, self-reported use captured via structured household use schedules, instrument-recorded use, and latrine spot check indicators) compare?

2. METHODS

This verification was conducted during June-August 2014, concurrent with WASH II project implementation⁴, in communities touched by BRAC's WASH II project. Statistical analyses were performed in Stata 13 (Statacorp, College Station, TX). The verification was approved by Emory University's Institutional Review Board as well as icddr,b's Ethical Review and Research Review Committees. Oral informed consent was obtained in Bangla, and was obtained for all verification participants.

2.1 Verification design

In collaboration with icddr,b, verification staff employed a post-only non-equivalent comparison group study design to obtain latrine coverage and use data via a semi-structured survey, direct observation via latrine spot checks, and electronic PLUMs. Our verification of BRAC's data included a comparative analysis of latrine coverage and utilization proportions reported in BRAC's narrative progress report as well as a direct comparison of QIS ladder scale score distributions between BRAC's QIS monitoring dataset and our verification dataset for all three of BRAC's QIS latrine coverage and use indicators. We compared BRAC's QIS data to sanitation outcome measures obtained via our administration of BRAC' QIS ladder scaling system, survey, and observation data.

In order to ascertain whether BRAC's QIS sample, selected from 299 Village WASH Committees (VWCs) within 178 unions (165 WASH I, 13 WASH II) and 72 upazilas (sub-districts - 69 WASH I, 3 WASH II), is representative of its larger project population, we surveyed a comparison group selected from VWCs within the 105 upazilas not included in BRAC's QIS sample. We selected an approximately similar number of households from QIS (VWCs within BRAC's QIS sample) and non-QIS VWCs (comparison VWCs from the larger WASH II project population, but outside of BRAC's QIS sample). A multi-stage cluster sampling strategy was employed to draw an appropriate verification sample.

2.1.1 Target verification population

BRAC categorizes all participant household wealth classifications as ultra-poor, poor, and non-poor via facilitating participatory rural appraisals (PRAs) led by VWC members, and targets all households within upazilas engaged under its WASH II project with indirect, "software" interventions (i.e., BCC, household and community consultation). In addition to these "software" interventions, they reportedly target 50% of all ultra-poor households in the project catchment area for direct hardware interventions (i.e., distribution of in-kind subsidies or loans for latrine installation/repair). Therefore, all households situated within the WASH II project upazilas comprised the targeted verification population.

2.1.2 Sampling frame

The enumerated list of VWCs monitored by BRAC through its QIS served as the sampling frame for our QIS group sub-sample. The enumerated list of all VWCs not included in BRAC's QIS sample served as the sampling frame for the comparison group sub-sample.

2.1.3 Primary and secondary sampling units

In order to obtain their QIS sample, BRAC reportedly sampled equally from WASH I-engaged communities and WASH II-engaged communities. Since WASH I-engaged communities reside in 152 upazilas, and WASH II-engaged communities reside only in 25 upazilas, BRAC selected two different primary sampling units for

⁴ The WASH II project commenced in October 2011, and BRAC conducted its QIS survey during 2013. We conducted the verification exercise a little over two and a half years after the project commenced.

each group - upazila and union for communities engaged under WASH I and WASH II, respectively. Our study used a similar approach to sampling WASH I and WASH II-engaged communities. As such, the primary sample unit for this verification is the upazila and union for communities engaged under WASH I and WASH II, respectively. The union and VWC were secondary sampling units for WASH I and WASH II, respectively. VWCs are the lowest organizational level within BRAC's WASH II project, and the size of each VWC is relatively similar (there are typically 100-300 households per VWC catchment area).

After receiving BRAC's QIS dataset, we came to learn that the QIS sample was not selected equally from WASH I and WASH II-engaged communities, as reported. Instead, 87.3% (6536/7489) of the sample was selected from WASH I communities, and 12.7% (953/7489) of the sample was selected from WASH II communities. We determined this information prior to the initiation of the data collection phase, and asked the Foundation how they would like to move forward. The decision was made to sample equally from WASH I and WASH II communities, as previously planned.

2.1.4 Ultimate sampling unit

The ultimate sampling unit for this study is the household; specifically, any household residing in an upazila targeted by BRAC's WASH II project.

2.1.5 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 a Monte Carlo simulation in SAS to repeatedly simulate the outcomes of interest from model parameters estimated from BRAC's QIS data. 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.

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. Using the largest required sample size allowed us to be adequately powered for all indicators of interest. Our power analysis indicated that 52 clusters (with an 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 accounted for the clustered nature of the data. Given 9% (26 of 299) of VWC clusters surveyed by BRAC contained less than our targeted 21 households, and 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 in each VWC (for a total targeted sample size of 1,248 households). Further details regarding our sample size calculation and justification of our final sample size can be found in Appendix I.

2.1.6 Sample selection

Verification staff employed a multi-stage sampling strategy to select the 24 households per selected cluster. During the first stage, we selected 26 PSUs from WASH I-engaged communities (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-

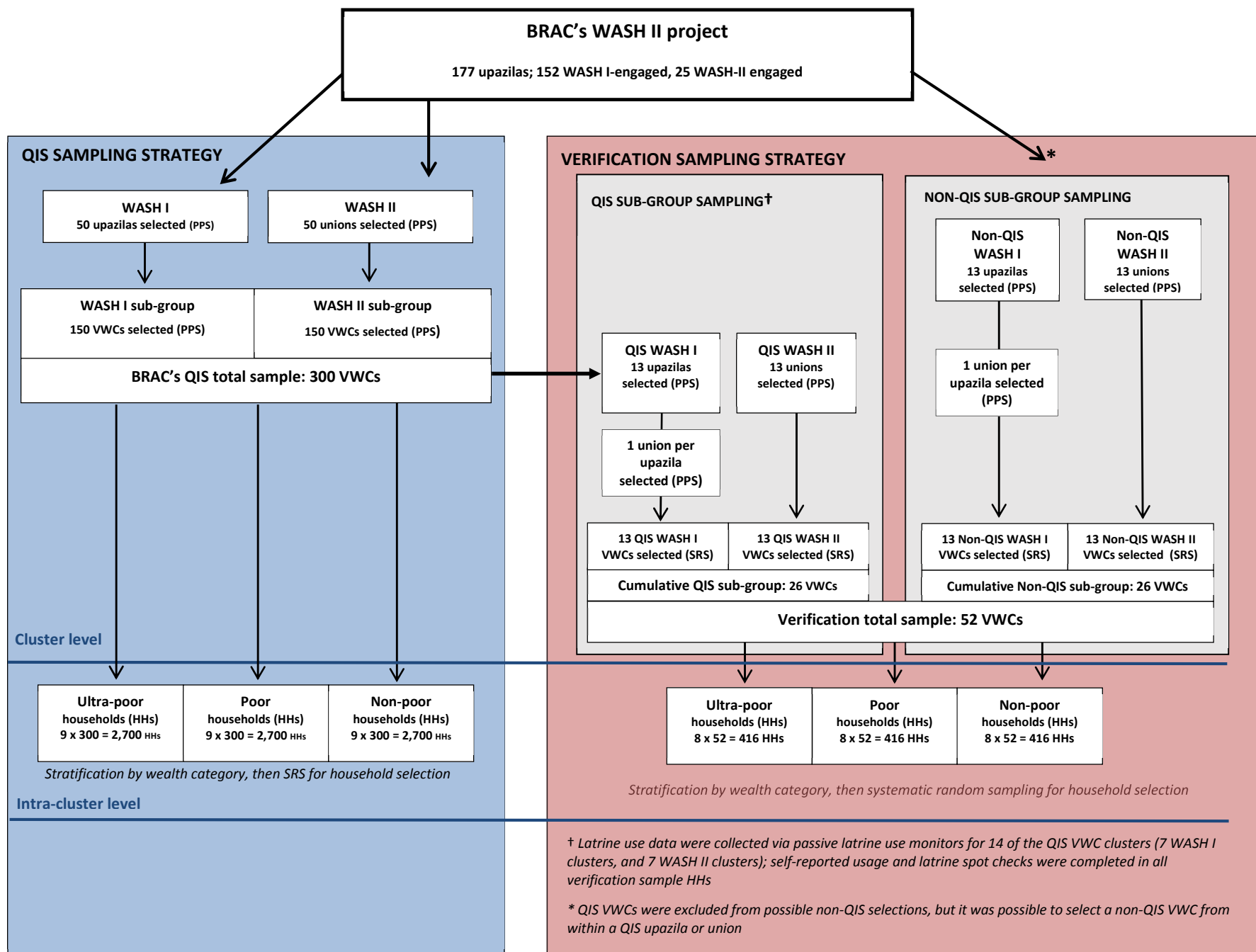
engaged communities (we selected 13 unions from the QIS sampling frame to help comprise our QIS sub-group, and 13 unions from the non-QIS sampling frame to help comprise our non-QIS group), using probability proportional to size (PPS). During the second stage in WASH I communities, one union was selected per upazila, via PPS. During the next sampling stage for WASH I and WASH II communities, one VWC was selected from each union via simple random sampling (SRS). In the final stage, we obtained the VWC register (household sampling frame) from selected VWCs, and stratified the sampling frame by wealth category as defined by BRAC/VWC (i.e., ultra-poor, poor, non-poor). Finally, 8 households were selected per wealth category via systematic random sampling (i.e., households were enumerated by wealth category; the sampling interval [k] was calculated by dividing the total number of households in the wealth category by 8; a random number between 1 and k was generated to identify the first household for selection into the sample; and the sampling interval was applied from that point forward to select the sample for each wealth category). In order to allow for random replacement of households in the field,⁵ verification staff generated a list of 15 randomly selected households per wealth category. During survey administration, enumerators targeted the first 8 households per wealth category in each VWC. If one of those 8 households was vacant, had no eligible survey respondent present to represent the household, or refused to participate in the survey, the enumerator replaced the household with the ninth household on the list, and so on, until a total of 8 households per wealth category were surveyed in each selected VWC. Field supervisors were on hand to coordinate household replacement. Figure 4 outlines the sampling strategy employed by BRAC (per their QIS documentation, though, as indicated above, there was a deviation from the documented strategy) and the verification team.

Inclusion criteria: All households residing within selected VWCs in BRAC's WASH II project upazilas were eligible for inclusion in the study 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 (including poor and ultra-poor) project interventions. In order for households to be selected for inclusion in the QIS group sub-sample, the household had to reside in a VWC that was included in BRAC's QIS survey. In order for households to be selected for inclusion in the comparison group sub-sample, the household had to reside within one of the upazilas not included in BRAC's QIS sample. In order to be eligible to participate in the survey, households had to have at least one adult who consented to participate in the verification, and serve as the primary survey respondent. Further, the adult individual also had to be capable of understanding and providing informed consent.

Exclusion criteria: Any household selected for our study sample that refused to be surveyed, or either was repeatedly vacant after three attempts or did not have an appropriate member (i.e., capable female/male 18 years or older) of the household home to serve as the 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.

⁵ Enumerators were trained to visit the first 8 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 refused to participate in the study were replaced by the next randomly selected household on the list.

Figure 4. Sampling strategy – QIS and verification samples



2.2 Outcome measurement

Un-weighted proportions are presented throughout the report, as direct project interventions are implemented at the household level, and these proportions provide the reader with an understanding of sample observations. Population-averaged proportions are also presented, and specified as such, throughout the text, in numerous figures, and the accompanying data tables. These population-averaged proportions adjust for study design, correct for block sampling amongst the three wealth categories at the VWC level, and provide the reader with estimates of the proportion of the *overall verification population* maintaining the characteristics of interest. We used the `svy tabulate` command in Stata to generate these population-averaged descriptive statistics.

Individuals within groups or clusters may have correlated outcomes due to similar group context, interaction between individuals (e.g., information sharing within groups), and/or common intervention experiences within each group. In other words, individuals within groups or clusters tend to be more similar to each other than to individuals in other groups or clusters. As an artifact of our clustered sampling design, our data are correlated at the VWC cluster level, and therefore violate independence assumptions made by traditional regression procedures. This violation is most relevant to the estimates of the variability of mean population-averaged point estimates (i.e., correlation influences standard errors accompanying mean point estimates [not the mean estimates themselves], as these unadjusted standard errors are typically too small [i.e., they result in inflated Type I errors]) (Hubbard et al. 2010). If traditional regression approaches are used for hypothesis testing, they may generate evidence that suggests falsely rejecting null hypotheses.

Consequently, we determined the need to use an analytical regression approach that accounted for correlated outcome data. Since we were seeking to obtain population-averaged models whose parameters estimate the marginal population mean, we decided to use generalized estimating equations (GEE) with a working exchangeable correlation matrix in the standard error calculation to analyze data for our comparative analyses. We also decided to use a robust standard error model specification that uses the observed variability in the data (as opposed to the variability predicted by the underlying probability model) and household-level residuals to take intra-cluster correlations into account. P-values generated from Wald hypotheses tests are presented for hypothesis test results. One p-value, generated via the `testparm` post-estimation command, is presented for multi-level covariates. Small p-values indicate evidence against the null hypothesis that the two sample means under comparison are equal.

2.2.1 Latrine coverage

During the verification exercise, latrine coverage was assessed through two different methodologies: 1. BRAC's QIS ladder scaling, in which a household's access to a hygienic latrine was scored according to set criteria; and 2. surveying of household latrines, in which enumerators first asked respondents about access to household latrines, and then directly observed each latrine. During direct observation of each latrine, enumerators assessed aspects of latrine construction and functionality.

BRAC's QIS latrine coverage indicator (HH03)

In order to ensure comparability of data for our comparative analyses, we incorporated a separate module in our household questionnaire in which we administered the three relevant QIS survey prompts. To minimize survey administration bias, we invited trainers from BRAC to train icddr,b enumerators on how to administer the QIS survey prompts and how to scale household scenarios via BRAC's QIS "ladders". In order to further limit administration bias, particularly on the QIS questionnaire module (i.e., enumerators adopting a more systematic strategy for inquiring about latrine utilization for all members of the household), we designated specific enumerators to administer the QIS module only. We selected our

female enumerators to administer the QIS module, as it is culturally more appropriate for women to speak more openly with other women than men. In more conservative settings, these QIS module enumerators were also required to administer the remaining modules in household questionnaire, as the enumerators who administered these modules were mostly men. This cross-over potentially introduced bias into the administration of the QIS module.

We generated cluster-adjusted risk differences using generalized estimating equations, with standard errors adjusted for clustering on the VWC level. Risk differences provide an absolute measure of association between an exposure (i.e., the BRAC WASH project, in this case) and an outcome of interest (i.e., household latrine coverage and utilization). An absolute measure is important when cost-effectiveness is being considered. We calculated risk differences between BRAC's QIS sample and the total verification sample as well as BRAC's QIS sample and the verification QIS sub-group and BRAC's sample and the verification non-QIS sub-group by reported sanitation coverage and utilization scale on a dichotomous level (i.e., at/above or below threshold, as per BRAC's narrative progress report) and on a binomial level for each step of BRAC's QIS sanitation ladder. Through the use of generalized estimating equations, we chose to fit separate logistic regressions for each level of the QIS ladder (i.e., A-E, A-8, and A-E for HH03, HH04, and HH05 indicators; respectively). This allowed for a closer investigation of each step of the QIS ladder. Although proportional odds models (in which each dichotomization of each level of the QIS sanitation ladder would be assessed over all possible ladder scores) provide a more statistically efficient method for ordinal logistic regressions, these two different methods produce the same approximation of effect estimates.

Household questionnaire

At selected households, enumerators sought out adult respondents, with preference going first to the primary female caretaker of the youngest child within the household (as she will tend to know the most about the latrine use and defecation practices of most members of her household). If she was not available, they sought out other household members in the following order: 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.

Emory and icddr,b staff developed a survey instrument consisting of several different modules that aimed to collect information from respondents regarding:

- their basic demographics and socioeconomic status⁶;
- latrine construction, maintenance, and repair;
- latrine structure and functionality;
- latrine spot checks;
- reported latrine use of each household member,⁷ including use for disposal of child feces;

⁶ In order to ensure comparability between our survey data and BRAC's QIS survey data, we structured our survey questions around the indicators BRAC uses to define ultra-poor, poor, and non-poor households. We structured additional questions around the Government of Bangladesh (GoB) definitions for these wealth categories. While we will use the VWC's definition of household wealth status to select our study sample (as did BRAC for its QIS sample), we will obtain household level, self-reported data to assess the accuracy of those classifications. In our data analyses, we compared our wealth status classifications with BRAC's QIS survey classifications. Additionally, we compared our wealth status classifications using BRAC's definitions to our wealth status classifications using GoB definitions to assess comparability.

⁷ The primary household survey respondent reported on behalf of each member of the household. If other household members are present during survey administration, they were asked to self-report on their own latrine use behaviors.

- reported non-hardware project intervention exposure (e.g., exposure to and acceptance of BRAC behavior change communication messages related to sanitation and hygiene, community mobilization events);
- distances between the latrine and various points of interest (e.g., the homestead, latrine water source);
- the three sanitation indicators developed by BRAC via their QIS; and
- recent illness/symptom reports (e.g., diarrhea/dysentery and respiratory symptoms indicative of acute respiratory infection [ARI]) for every member of the household.

Enumerators administered all survey modules at all households in the sample, such that all households in the verification sample provided self-reported latrine use data, and all household latrines in the verification sample underwent a latrine spot check (as the time it would take to collect latrine spot check data was minimal given the need for enumerators to go to the latrines to visually observe, verify, and record various aspects of latrine construction and functionality). Enumerators collected survey data on password-protected, Samsung Galaxy 3 tablets. The electronic survey entry form included programming that allowed for range and consistency checks. In order to record and present costing information for each latrine use measurement method, enumerators recorded start and end times for each survey module. Person-time and other relevant costs were incorporated into each measurement method's costing.

Direct observation

After inquiring about reported access and functionality of household latrines, enumerators requested survey respondents to show them each of the latrines, where the enumerators assessed each latrine to determine the type of facility, and assess various attributes of latrine structure and function.⁸ During this process, the enumerator asked the survey respondent to provide information related to the dates of each latrine's original construction and most recent repairs (while the enumerator and respondent were physically standing at the latrine, so as to avoid confusion regarding which particular latrine was being discussed), and inquired whether, by whom, and in what form (i.e., subsidy, loan) assistance was provided to construct and/or repair the latrine. After ascertaining pertinent information related to the latrine, the enumerator conducted a latrine spot check (see details in the section 2.2.2 sub-section entitled *Latrine spot check indicators via direct observation*).

2.2.2 Latrine utilization

Evidence of open defecation

In addition to surveying households regarding their latrine utilization habits, verification team members recorded whether they observed evidence of open defecation (i.e., presence of human feces) in each surveyed VWC. Verification staff made note of whether evidence of open defecation (OD) was observed at the household level (i.e., within household compounds), at the community level (i.e., within open spaces in the VWC), or at both household and community levels. These data were captured in order to triangulate self-reported and instrument-recorded utilization data.

BRAC's QIS latrine use indicators: "latrine use, by whom" (HH04); "latrine use, when" (HH05)

Similar methods were used to calculate risk differences between BRAC's QIS sample and the total verification sample as well as BRAC's QIS sample and the verification QIS sub-group and BRAC's QIS sample and the verification non-QIS sub-group for BRAC's latrine use indicators.

⁸ Enumerators made assessments of the following latrine attributes: presence, height, and material of latrine enclosure; material of latrine door, roof, and floor; number of latrine pits and rings; condition of the latrine pan; and status of latrine water seal, latrine pit cover, and latrine pit/tank.

Self-reported latrine utilization via structured household use schedule

Enumerators administered a structured household latrine use schedule which systematically captured data on latrine use for each regular user of the household latrine(s), including household members and other regular users (e.g., neighbors, tenants, servants).

Individual-level reported use

For each household latrine member, respondents (or actual latrine users, if present) were asked about the primary place of defecation, whether the primary place of defecation changes during the year, and whether latrine users always exclusively use the latrine for defecation. Respondents were subsequently queried on the number of times each household latrine member uses the latrine during four specific periods of the day— morning (i.e., 04:00-10:00), afternoon (i.e., 10:00-15:00), evening (i.e., 15:00-19:00), and night (i.e., 19:00-04:00). Based on the manner in which these survey prompts were translated and administered by enumerators, we feel that although the specific hours of the day may not always have been relevant to local populations, they were still able to distinguish between these different times of day (e.g., enumerators asked respondents how many times they used the latrine during the early part of the day, then indicated the exact hours, and continued with this approach for mid-day, evening, and night-time). While we believe people were able to distinguish between the various times of day, these distinctions in time of day are less relevant to our outcome measure, as the primary purpose of obtaining latrine use data over four periods of the day was to collect data in a structured manner so events were less likely to be overlooked. During the analytical phase, we created a variable that generated the total number of reported daily latrine events for each household. For our comparative analysis of self-reported versus instrument-recorded use, we created a variable that generated the product of the household daily total by four days, so as to allow us to compare utilization between these two measurement methods on the same scale. For reporting purposes, we also calculated the average number of self-reported latrine events per household and per person for each of the four daily reporting periods (i.e., morning, afternoon, evening, night).

Although the language on the three survey prompts preceding the daily latrine event capture was specific to defecation behavior, the language on the survey prompts quantifying latrine use was more general, and inquired about latrine use as opposed to latrine use for defecation purposes. Given it is customary for latrines to be used primarily for defecation in rural areas, this was not believed to have largely biased the self-reported latrine use estimates. However, this lack of specificity in the survey prompts did likely introduce some level of reporting bias. It is important to make note of this, as the validated algorithm used to analyze PLUM signal data is designed to detect defecation events (as opposed to latrine events). Analyses comparing self-reported and instrument-recorded utilization therefore compare self-reported latrine events with PLUM-recorded defecation events.

Household-level use classifications

In terms of classifying household-level latrine utilization, we assessed exclusive latrine use for defecation amongst all latrine users in the household. We assigned household-level defecation designations based on these responses. Households with all latrine members always exclusively defecating in a latrine were classified as “always using” households; households with mixed use patterns amongst its latrine users were classified as “sometimes using” households; and households with all latrine members never exclusively defecating in a latrine were classified as “never using” households. While we concurrently assessed whether each individual exclusively used a latrine for defecation along with each individual’s primary place of defecation, exclusive latrine use for defecation was the decisive variable in assigning these household-level latrine use designations when inconsistencies arose between the two data points.

Latrine spot check indicators via direct observation

After ascertaining pertinent information related to the latrine, the enumerator conducted a latrine spot check, which consisted of a structured observation during which the enumerator observed and recorded the presence or absence of a series of use indicators.⁹

Instrument-recorded use data via PLUM sensors

All consenting households within a randomly selected sub-set of QIS VWCs had a PLUM placed in all functional household latrines for one week, such that just over one-half of randomly selected VWCs in the QIS group sub-sample were included (i.e., 14 VWCs [7 WASH I and 7 WASH II VWCs]). We believed that targeting just over half of the QIS sub-group with PLUM deployment would allow us to actually obtain data from at least one-third of the QIS households 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. A specialized team of field staff were trained in PLUM installation and removal.

We worked with collaborators at Portland State University to analyze PLUM signal data collected over the one-week PLUM data collection period. It was felt that there may be considerable reactivity in the first several days after PLUM installation (e.g., curious children and adults entering the latrine to look at the sensor, people perhaps modifying their latrine use behaviors). Therefore, we spliced PLUM data such that the first two days of data collection, and the last one day of data collection were dropped. We used the remaining four days of PLUM data for our analytical sample. PSU collaborators used a validated algorithm that assesses raw PLUM signal data, and detects and quantifies signal patterns indicative of defecation events (Clasen et al. 2012).

2.2.3 Latrine utilization measurement method cost assessment

In order to compare the costs associated with the three main latrine utilization measurement methods of interest (i.e., self-reported use, instrument-recorded use, and latrine spot check indicators of use), we calculated the amount of time required to administer each associated survey module. For PLUM costing, there was no associated survey module; therefore, we calculated the amount of time required to install and remove each PLUM deployed to household latrines. Once we calculated the total amount of minutes required to administer the respective survey modules and the amount of time required for PLUM installation/removal, we converted the figure to person-hours. Since the number of PLUMs deployed was quite a bit lower than the number of households from which we obtained self-reported use and spot check use indicator data, we calculated the average administration time per unit. In order to obtain the cost associated with administering the latrine utilization measurement method, we multiplied the average time per unit by the hourly rate for one fixed-term Field Research Assistant. On the advice of PSU, we used a value of \$50 per month per unit as the cost of renting the PLUMs, providing technical support, and conducting the analysis for converting the signal data into defecation events. Similarly, we included a monthly rental rate for handheld tablet devices, as data from household use schedule surveys and latrine spot checks were electronically entered by enumerators, and this type of data entry may well affect the amount of time spent administering respective survey modules. A value of \$30 per month per unit was derived from the purchase price of the tablets in 2013, with an assumed product life of 24 months.

⁹ Spot check indicators included: evidence the latrine is used for a purpose other than urination/defecation, stagnant water over the latrine floor/slab, visible traces of feces in/on the pan/slab, discoloration of pan/slab, odor of stool/urine, presence of debris in the latrine pit/pan, presence of flies in the latrine, availability of anal cleansing agents, evidence of a well-worn path to the latrine, wet floor, availability of water and cleansing agents for handwashing, and presence of slippers near the latrine.

The purpose of this cost assessment was to compare costs associated with the three latrine utilization methods; therefore, we did not include other field-related costs such as those associated with field per diems, car rental, fuel, accommodation, training-related cost, field supervisory and support staff, and other miscellaneous expenses. Our underlying assumption was that these costs would be incurred uniformly regardless of the method used to measure latrine utilization. Total costs were not presented, as this was not intended to be an accounting exercise, and so the reader can apply these estimated costs to the number of units and the amount of months necessary for future verifications.

3. RESULTS

Sample characteristics

The verification sample includes data from households residing in 52 VWCs within 49 unions from 30 distinct upazilas in Bangladesh. The sample is comprised almost equally of WASH I and WASH II households (52% [631/1207] vs. 48% [576/1207] for WASH I and WASH II households, respectively), and QIS and non-QIS households (50% [607/1207] vs. 50% [600/1207] for QIS and non-QIS households, respectively). BRAC's QIS sample includes data from households residing in 299 VWCs within 178 unions (165 WASH I, 13 WASH II) from 72 distinct upazilas in Bangladesh. BRAC's analytical sample is drawn primarily from WASH I households (87% [6536/7489]).

As indicated in Table 1, females were the primary survey respondents in both BRAC QIS and verification samples. The mean household size was slightly larger for BRAC's QIS sample compared to the verification sample (5.2 persons [interquartile range (IQR) 4, 6] vs. 4.7 persons [IQR 4, 6], respectively). After adjusting for block sampling, the samples were similar with regard to wealth status distributions. Additional details regarding respondent demographics, household demographics, and socio-economic indicators for the verification sample can be found in Tables A1a and A1b in Appendix II.

Table 1. Respondent characteristics for QIS and verification survey samples

Characteristic	Households (un-weighted)		Population (weighted)			
	BRAC's QIS sample	Verification sample	BRAC's QIS sample		Verification sample	
	n (%)	n (%)	n (%)	95% CI	n (%)	95% CI
Sex of main respondent						
Female	6,231 (83.2)	1,073 (88.9)	48,670 (83.0)	(42210, 55131)	10,278 (88.9)	(8874, 11682)
Household wealth category*						
Non-poor	2,555 (34.1)	416 (34.5)	29,052 (49.5)	(24367, 33737)	5,143 (44.5)	(4199, 6087)
Poor	2,477 (33.1)	381 (31.6)	17,959 (30.6)	(15715, 20203)	3,346 (28.9)	(2425, 4267)
Ultra-poor	2,457 (32.8)	410 (34.0)	11,646 (19.9)	(10024, 13268)	3,078 (26.6)	(2588, 3568)
Household size	mean (IQR)	mean (IQR)	mean	IQR	mean	IQR
Mean household size	5.2 (4, 6)	4.7 (4, 6)	5.4	(4, 6)	4.7	(4, 6)
Number	7,489	1,207	58,657	--	11,567	--

Notes:

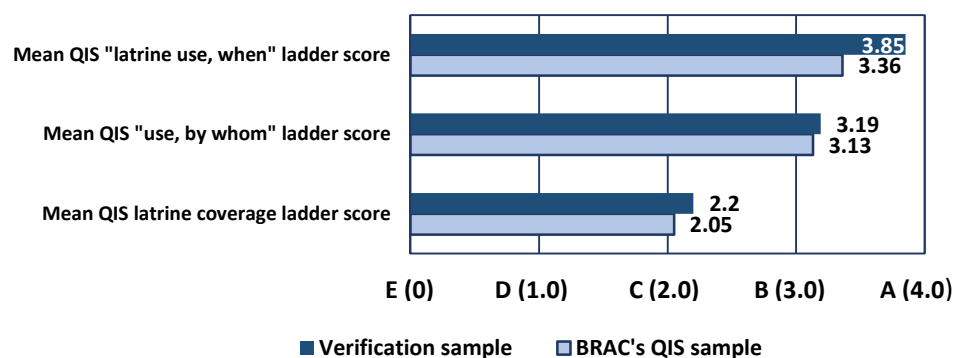
IQR: interquartile range; CI: confidence interval

* Household wealth category per BRAC's 2007/2012 census, per VWC register

3.1 VERIFICATION QUESTION 1: Are BRAC's monitoring data accurately reflective of latrine coverage and use among WASH project participants?

Figure 5 provides population-averaged summary estimates of mean QIS ladder scores for the three QIS sanitation outcome indicators in surveyed areas. Specific details regarding the comparability of each sanitation outcome indicator are indicated under relevant results sections below.

Figure 5. Population-averaged mean QIS ladder score estimates for BRAC QIS and verification samples



3.1.1 Latrine coverage

3.1.1.1 QIS latrine coverage indicator

Through its QIS system, BRAC assesses latrine coverage by assessing access to a hygienic latrine, and assigns the household a score according to a standardize scale for latrine coverage. For comparative purposes, assessments against narrative progress reports follow the specifications noted in the narrative progress report (e.g., hygienic latrine coverage at or above benchmark was assessed amongst ultra-poor and poor households [as an aggregate] and non-poor households [separately], as these are the levels BRAC reported in its narrative progress report).

QIS latrine coverage indicator

Although there were some differences in QIS score distributions along the latrine coverage ladder, score distributions were similar between BRAC's sample and the verification sample. Internal consistency was observed in the verification sample in that similar deviations from BRAC's QIS ladder score distributions were observed in the total verification and the verification QIS/non-QIS sub-groups. For example, where there was a significant difference between BRAC's QIS sample and the total verification sample, for the most part, there was also a significant difference between BRAC's QIS sample and the verification QIS or non-QIS sub-groups. In assessing reported latrine coverage benchmarks, as indicated by BRAC in their narrative progress report, no significant differences were observed in reported latrine coverage at or above benchmark amongst non-poor households. However, a significantly higher proportion of verification QIS sub-group households were at or above benchmark compared to BRAC's QIS sample (67.7 vs. 57.0%, respectively; Risk Difference [RD]=9.7, 95% CI 0.8, 18.5).

Further details regarding QIS latrine coverage ladder score and benchmark distributions are presented in Table 2. Readers should note that data presented in the first five rows of the table represent a direct comparison between the latrine coverage ladder scores between BRAC's QIS dataset and our verification dataset, while the data presented in the last two rows represent a verification of BRAC's reported results, as per the wording used in their narrative progress report. Table 2a displays results with non-poor and ultra-poor/poor disaggregations for each level of the latrine coverage ladder scale. This table indicates some effect modification between non-poor vs. poor and ultra-poor households.

Table 2. QIS latrine coverage indicator HH03: comparison between BRAC's QIS sample and the verification sample, and BRAC QIS sample and verification QIS and non-QIS sub-samples

Indicators	BRAC's QIS sample	Total verification sample	RD [†] between BRAC QIS and verification samples	Verification QIS sub-sample	RD [†] between BRAC QIS and verification QIS samples	Verification non-QIS sub-sample	RD [†] between BRAC QIS and verification non-QIS samples
	% (n/N)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)
QIS latrine coverage indicator HH03:							
Score "A": Household had a latrine with: Ring and slab + functional water seal + absence of visible feces in pan, slab, water seal, and walls + latrine with two pits	7.7 (579/7489)	13.8 (167/1207)	6.3 (2.9, 9.7)*	16.0 (97/607)	8.2 (3.0, 13.3)*	11.7 (70/600)	4.4 (0.6, 8.2)*
Score "B": Household had a latrine with: Ring(s) and slab + functional water seal + absence of visible feces in pan, slab, water seal, and walls	35.0 (2624/7489)	24.4 (294/1207)	-10.6 (-15.0, -6.2)*	28.0 (170/607)	-6.9 (-12.5, -1.2)*	20.7 (124/600)	-14.4 (-19.4, -9.4)*
Score "C": Household had a latrine with: Ring(s) and slab+ functional water seal	19.1 (1429/7489)	28.0 (338/1207)	8.5 (4.5, 12.6)*	26.7 (162/607)	7.2 (1.5, 13.0)*	29.3 (176/600)	9.9 (4.9, 14.8)*
Score "D": Household had a latrine with: Ring(s) and slab (no or broken water seal)	26.6 (1988/7489)	31.5 (380/1207)	5.2 (-1.0, 11.3)	27.0 (164/607)	0.8 (-6.6, 8.2)	36.0 (216/600)	9.5 (0.6, 18.5)*
Score "E": Household had a latrine with: No latrine or latrine without ring or slab	11.6 (869/7489)	2.3 (28/1207)	-9.3 (-11.6, -7.1)*	2.3 (14/607)	-9.3 (-11.7, -6.9)*	2.3 (14/600)	-9.4 (-11.8, -6.9)*
Assessing against narrative progress report: At or above the latrine coverage benchmark (where benchmark refers to a latrine with (1) rings and a slab, and (2) a functioning water seal – per BRAC's definition) AMONG ULTRA-POOR <u>AND</u> POOR HOUSEHOLDS	57.0 (2811/4934)	62.5 (494/791)	4.6 (-2.7, 11.9)	67.7 (270/399)	9.7 (0.8, 18.5)*	57.1 (224/392)	-0.5 (-10.5, 9.4)
Assessing against narrative progress report: At or above the latrine coverage benchmark (where benchmark refers to a latrine with (1) rings and a slab, and (2) a functioning water seal – per BRAC's definition) AMONG NON-POOR HOUSEHOLDS	71.3 (1821/2555)	73.3 (305/416)	2.1 (-4.9, 9.0)	76.4 (159/208)	5.2 (-3.3, 13.7)	70.2 (146/208)	-1.1 (-10.9, 8.8)

[†]RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VVC level

*Statistically significant at $\alpha=0.05$ ($p<0.05$)

Table 2a. QIS latrine coverage indicator HH03: comparison between BRAC's QIS sample and the verification sample, and disaggregation between non-poor and ultra-poor/poor HHs

Indicators	BRAC's QIS sample		Total verification sample		RD [†] between BRAC QIS and verification samples		Verification QIS sub-sample		RD [†] between BRAC QIS and verification QIS samples		Verification non-QIS sub-sample		RD [†] between BRAC QIS and verification non-QIS samples	
	% (n/N)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)	
	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs
QIS latrine coverage indicator HH03:														
Score "A": Household had a latrine with: Ring and slab + functional water seal + absence of visible feces in pan, slab, water seal, and walls + latrine with two pits	5.1 (130/2555)	9.1 (449/4934)	17.6 (73/416)	11.9 (94/791)	12.4 (7.3, 17.5)*	3.0 (-0.5, 6.5)	19.7 (41/208)	14.0 (56/399)	14.6 (7.5, 21.7)*	5.0 (-0.4, 10.3)	15.4 (32/208)	9.7 (38/392)	10.2 (3.3, 17.2)*	1.0 (-2.6, 4.6)
Score "B": Household had a latrine with: Ring(s) and slab + functional water seal + absence of visible feces in pan, slab, water seal, and walls	46.5 (1187/2555)	29.1 (1437/4934)	30.5 (127/416)	21.1 (167/791)	-15.7 (-21.9, -9.5)*	-8.3 (-13.3, -3.3)*	35.1 (73/208)	24.3 (97/399)	-11.1 (-19.1, -3.2)*	-5.1 (-11.9, 1.6)	26.0 (54/208)	17.9 (70/392)	-20.3 (-28.1, -12.5)*	-11.5 (-17.4, -5.7)*
Score "C": Household had a latrine with: Ring(s) and slab + functional water seal	19.7 (504/2555)	18.8 (925/4934)	25.2 (105/416)	29.5 (233/791)	5.4 (-0.50, 0.11)	10.1 (5.7, 14.6)*	21.6 (45/208)	29.3 (117/399)	1.8 (-6.5, 10.1)	10.1 (3.7, 16.4)	28.9 (60/208)	29.6 (116/392)	9.0 (1.4, 16.6)*	10.2 (4.7, 15.7)*
Score "D": Household had a latrine with: Ring(s) and slab (no or broken water seal)	21.0 (536/2555)	29.4 (1452/4934)	26.0 (108/416)	34.4 (272/791)	5.0 (-1.6, 11.6)	5.7 (-1.1, 12.6)	23.6 (49/208)	28.8 (115/399)	2.6 (-5.7, 10.9)	0.2 (-8.2, 8.5)	28.4 (59/208)	40.1 (157/392)	7.4 (-2.2, 17.0)	11.3 (1.7, 21.0)*
Score "E": Household had a latrine with: No latrine or latrine without ring or slab	7.8 (198/2555)	13.6 (671/4934)	0.7 (3/416)	3.2 (25/791)	-7.0 (-9.1, -4.9)*	-10.4 (-13.1, -7.8)*	0.0 (0/208)	3.5 (14/399)	-	10.0 (-12.9, -7.1)*	1.4 (3/208)	2.8 (11/392)	-6.3 (-8.8, -3.8)*	-10.9 (-13.8, -8.0)*

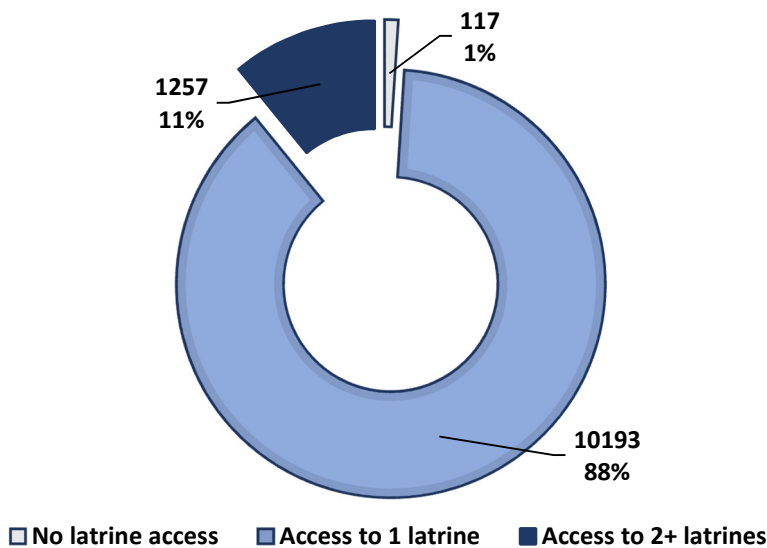
[†]RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VVC level

*Statistically significant at $\alpha=0.05$ ($p<0.05$)

3.1.1.2 Enumerator-assessed latrine coverage

Of the 1,207 verification households surveyed, 98.9% (1194/1207) reported access to at least one household latrine. These 1,207 households have access to a total of 1,328 household latrines (i.e., 116 households have access to more than one latrine). The majority of verification households have access to only one household latrine (89.3% (1078/1207), but 9.6% (116 households) have access to two or more household latrines (105 households have access to two latrines, six households have access to three latrines, three households have access to four latrines and two households have access to five latrines). Figure 6 depicts population-averaged household latrine access for areas sampled during the verification. There is marginal evidence that access to any household latrine amongst verification households is associated with wealth category ($p=0.052$), as poor and non-poor households are 1.02 (95% CI 1.00, 1.04) times as likely to have access to a latrine.

Figure 6. Population-averaged estimates of household latrine access



Enumerator-assessed functionality

Of the 1,328 latrines surveyed at verification households, enumerators determined 98.6% (1310/1328) were functional per a list of pre-determined, standardized functionality criteria. The vast majority (99.7% [1075/1078]) of households with access to only one latrine have a functional latrine. Although enumerators deemed only 88.8% (119/134) of all secondary household latrines functional, all 116

households with non-functional secondary latrines have access to a functional primary latrine. Access to any type of functional latrine amongst verification households is not significantly associated with wealth status ($p=0.595$).

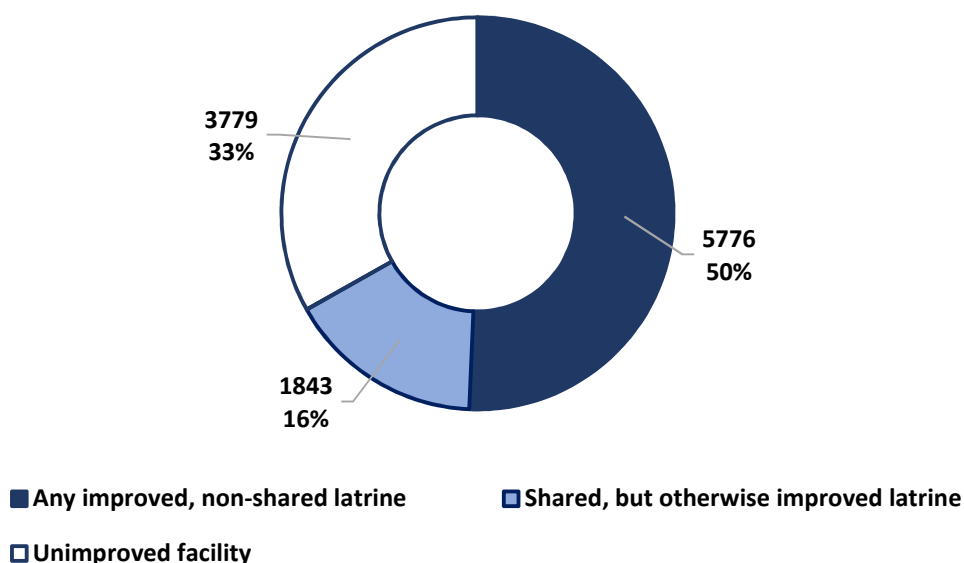
Type of sanitation facility

In assessing the type of sanitation facility(ies) to which verification households have access, we used the World Health Organization (WHO) and the United Nations Children's Fund's (UNICEF) Joint Monitoring Program (JMP) definitions for improved sanitation facilities. While JMP considers any shared latrine to be unimproved, we presented latrine type on three levels: improved¹⁰ (per JMP definition), shared facilities of an otherwise improved type (i.e., latrines that would be considered improved per JMP definition had they not been used by more than one household), and unimproved. Amongst the 1,207 verification households, 49.5% (597/1207) maintain ownership of one or more improved (not shared), functional latrines, 15.5% (187/1207) have access to one or more functional, shared sanitation facility of an

¹⁰ JMP considers the following to be improved sanitation facilities: flush or pour-flush to piped sewer system, septic tank, or pit latrine; a ventilated improved latrine; a pit latrine with slab; or composting toilet. JMP considers the following to be unimproved sanitation facilities: flush or pour-flush to elsewhere; pit latrine without slab/open pit; bucket; hanging latrine; shared facilities of any type; or no facilities (<http://www.wssinfo.org/>). For this analysis, we used

otherwise improved type, and 33.4% (403/1207) only have access to functional, unimproved latrines. In our verification sample, functional, unimproved facilities primarily consisted of open pit latrines (99.0%, 399/403), though a few hanging toilets/latrines were also observed (1.0%, 4/403). Just under one third (27.1%, 109/403) of these unimproved latrine facilities were also shared. Figure 7 presents population-averaged estimates of access to sanitation facility type in areas surveyed during the verification.

Figure 7. Population-averaged estimates of access to sanitation facility type



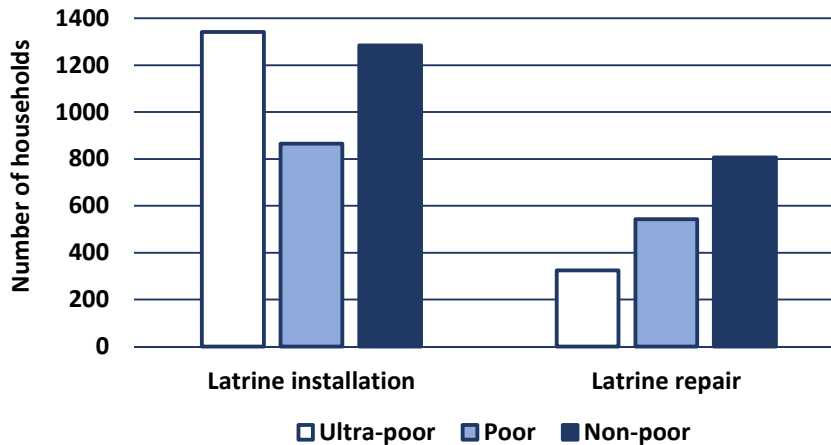
Household wealth status is strongly associated with access to only functional, unimproved household latrines ($p=0.001$), as ultra-poor and poor households are 1.14 (0.95, 1.37) and 1.37 (1.15, 1.63) times as likely as non-poor households to only have access to a functional, unimproved household latrine.

Self-reported latrine construction and repair

Amongst the 1,207 verification households, 41.8% (505/1207) have at least one improved or shared but otherwise improved latrine that was constructed or repaired during the WASH II project period. Of those 505 households, 337 (66.7%) have at least one improved or shared but otherwise improved latrine that was constructed (i.e., the type of latrine that is being promoted by BRAC's WASH project – exclusive of households reporting both construction *and* repair of a latrine), 128 (25.3%) households have at least one improved or shared but otherwise improved latrine that was repaired (exclusive of households that reported both construction *and* repair of a latrine), and 40 (7.9%) households have at least one improved or shared but improved latrine constructed *and* at least one improved or shared but improved latrine repaired during the WASH project period (i.e., the total number of households reporting construction of any improved or shared but otherwise improved latrine during the WASH II project period was 377, and the total number of households reporting repairs to any improved or shared but otherwise improved latrine was 168).

Households with improved or shared but otherwise improved latrines constructed during the WASH II project period represent 31.2% (377/1207) of the verification sample. There is a strong association between wealth status and construction of an improved or shared but otherwise improved latrine during BRAC's project period are associated ($p<0.001$), as poor and non-poor households were less likely than ultra-poor households to construct an improved or shared but otherwise improved latrine during BRAC's project period (Risk ratio [RR] =0.65 (0.54, 0.79) and RR=0.60 (0.47, 0.75), respectively. Figure 8 presents the population-averaged estimates of latrine installation and latrine repairs completed during the WASH II project period, by wealth category for areas surveyed during the verification.

Figure 8. Population-averaged estimates of latrine installation and latrine repairs completed during the WASH II project period, by wealth



The 377 verification households that constructed an improved or shared but otherwise improved latrine during BRAC's WASH II project reported constructing a total of 402 latrines. The 168 ultra-poor households reported a total of 170 improved or shared but otherwise improved household latrines were constructed during the project period, of which 166 households reported the construction of one latrine, and two households reported the construction of two latrines.

3.1.2 Latrine utilization

3.1.2.1 Evidence of open defecation (OD)

Some evidence of OD (i.e., human feces) was observed in 55.8% (29/52) of surveyed clusters. Amongst the 29 VWCs in which evidence of OD was observed, 51.7% (15/29) had evidence of OD in household compounds, 27.6% (8/29) had evidence of OD outside of household compounds in open spaces within the community, and 20.7% (6/29) had evidence of OD both inside household compounds and in open communal spaces. This evidence of OD should be considered along with latrine utilization results, as evidence of OD indicates latrine utilization for defecation is not exclusive. Table 3 outlines the level of evidence of open defecation for all 29 VWCs in which evidence of OD was observed.

During the verification exercise, latrine utilization was assessed through four different methodologies:

1. BRAC's QIS ladder scores, in which household latrine use is scored according to the individuals using the latrine (i.e., "latrine use, by whom" [HH04 indicator, per QIS]), and when these individuals use the latrine (i.e., "latrine use, when" [HH05 indicator, per QIS]);
2. self-reported use data captured via a household use schedule for all household latrine users;
3. instrument-recorded use data capturing defecation events during a four-day analytical period; and
4. latrine spot check indicators captured via direct observation of each household latrine.

Results are presented below, by latrine utilization measurement method.

3.1.2.2 QIS latrine use indicators

Through its QIS system, BRAC assesses latrine utilization on two levels:

1. a summarized assessment of latrine use amongst various categories of individuals in the household (i.e., latrine use, by whom), scored against a standardized scale of latrine use; and
2. a summarized assessment of when the latrine is used by these individuals (i.e., latrine use, when). For comparative purposes, assessments against narrative progress reports follow the specifications noted in the narrative progress report (e.g., "household members using" were assessed amongst ultra-poor households, as that is the level on which BRAC reported in its narrative progress report).

Table 3. List of surveyed VWCs with evidence of open defecation

SI #	Cluster #	VWC name	Union	Upazila	District	Place evidence of OD observed [†]
1	001	Atgaon-2	Atgaon	Bochagonj	Dinajpur	IC
2	002	Bolabunia	Batiaghata	Batiaghata	Khulna	IC
3	003	Chanpur	Dighirpar	Bajitpur	Kishoregonj	IC
4	005	Unchorokhi Dokhin	Gabtolli	Gabtolli	Bogra	Both
5	006	Bagutia-1	Ghagitia	Kapasia	Gazipur	IC
6	009	Gimadanga North Para	Patgati	Tungipara	Gopalganj	IC
7	010	Vaga	Rampal	Rampal	Bagerhat	CS
8	012	Hapania-1	Shodayse	Haluaghat	Mymensingh	Both
9	013	Musi Khali	Sonaray	Gabtolli	Bogra	IC
10	014	Chorpolash-2	Sukhia	Pakundia	Kishoregonj	IC
11	017	Adariatila	Datmara	Fatickchari	Chittagong	Both
12	020	Dowlotpur-2	Dowlotpur	Fatickchari	Chittagong	IC
13	021	Kazirchoura-1	Harati	Lalmonirhat Sadar	Lalmonirhat	IC
14	022	Teli Para	Kanchon Nagor	Fatickchari	Chittagong	CS
15	025	Chogram-1	Magura	Tala	Satkhira	IC
16	026	B. Fakirpara	Nanupur	Fatickchari	Chittagong	IC
17	051	Ambernagor-5	Ambernagor	Sonaimuri	Noakhali	CS
18	052	Kodim Hatil	Ausnara	Modhupur	Tangail	CS
19	054	Drugahata-3	Drugahata	Gabtolli	Bogra	IC
20	056	Simlapara-3	Hemnagor	Gopalpur	Tangail	CS
21	057	Tarakuri	Kailati	Netrakona Sadar	Netrakona	Both
22	058	Jotdoyboki-3	Lalpur	Lalpur	Natore	IC
23	060	Noajisipur-3	Noajisipur	Raozan	Chittagong	Both
24	061	Panisara-2	Panisara	Jhikargacha	Jessore	IC
25	063	A. Baid-1	Rasulpur	Ghatail	Tangail	CS
26	067	Tila Para	Bhujpur	Fatickchari	Chittagong	IC
27	068	Ramkrishnopur/Doripara	Dhalahar	Joypurhat	Joypurhat	CS
28	071	Muragacha-1	Khesra	Tala	Satkhira	CS
29	075	Nur Nagar Para	Nanupur	Fatickchari	Chittagong	Both

Notes: [†] Evidence of OD observed:

CS: Outside household compounds in communal spaces

IC: Inside household compounds

Both: Both inside household compounds and in communal spaces outside household compounds

QIS “latrine use, by whom” indicator (i.e., which people in the household use the household latrine)

As shown in Table 4, though there were some differences in QIS score distributions along the “latrine use, by whom” ladder, for the most part, score distributions were similar between BRAC’s sample and the verification sample. As with the latrine coverage QIS indicator, internal consistency was observed in the verification sample for the “latrine use, by whom” QIS indicator in that similar deviations from BRAC’s QIS ladder score distributions were observed in the total verification and the verification QIS/non-QIS sub-groups. In assessing the reported “latrine use, by whom” benchmark, as indicated by BRAC in their narrative progress report, a significantly higher proportion of ultra-poor verification households scored at or above benchmark compared to BRAC’s ultra-poor QIS households (87.6 vs. 75.4%, respectively; RD=12.1, [7.2, 16.9]). For comparisons between BRAC’s QIS sample and the verification QIS sub-group, and BRAC’s QIS sample and the verification non-QIS sub-group, the respective verification sub-groups had a higher proportion of households at or above benchmark compared to BRAC’s QIS sample for “latrine use, by whom” amongst ultra-poor households (89.1 vs. 75.4%, respectively; RD=13.7 [7.8, 19.6] and 86.1 vs. 75.4%, respectively RD=10.5 [4.3, 16.6]). Readers should note that data presented in the first six rows

of Table 4 represent a direct comparison between the latrine “use, by whom” ladder scores from BRAC’s QIS dataset against our verification dataset, while the data presented in the last row represents a verification of BRAC’s reported results, as per the wording used in their narrative progress report.

Table 4a presents results with non-poor and ultra-poor/poor disaggregations for each level of the latrine “use, by whom” ladder scale. This table indicates some effect modification between non-poor vs. poor and ultra-poor households.

Table 4. QIS “latrine use, by whom” indicator (HH04): comparison between BRAC’s QIS sample and the verification sample, and BRAC QIS sample and verification QIS and non-QIS sub-samples

QIS indicators	BRAC’s QIS sample	Total verification sample	RD† between BRAC QIS and verification samples	Verification QIS sub-sample	RD† between BRAC QIS and verification QIS samples	Verification non-QIS sub-sample	RD† between BRAC QIS and verification non-QIS samples
	% (n/N)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)
QIS “latrine use, by whom” indicator HH04 (use amongst those who have a latrine [hygienic or unhygienic]):							
Score “A”: Latrine used by: Women and adolescent girls + children ≥6 years age + men and adolescent boys + feces of other household members end up in toilet	41.8 (3131/7489)	37.5 (453/1207)	-3.9 (-9.2, 1.5)	36.9 (224/607)	-4.5 (-11.2, 2.2)	38.2 (229/600)	-3.3 (-9.7, 3.2)
Score “B”: Latrine used by: Women and adolescent girls + children ≥6 years age + men and adolescent boys	37.1 (2781/7489)	47.9 (578/1207)	10.3 (4.8, 15.8)*	47.5 (288/607)	9.8 (2.6, 17.0)*	48.3 (290/600)	10.8 (4.3, 17.3)*
Score “C”: Latrine used by: Women and adolescent girls + children ≥6 years age	8.2 (610/7489)	4.8 (58/1207)	-3.2 (-5.4, -1.1)*	6.1 (37/607)	-1.9 (-4.7, 0.9)	3.5 (21/600)	-4.6 (-6.9, -2.2)*
Score “D”: Latrine used by: Women and adolescent girls	8.1 (606/7489)	7.8 (94/1207)	-0.3 (-3.0, 2.4)	7.9 (48/607)	-0.2 (-4.0, 3.6)	7.7 (46/600)	-0.4 (-3.4, 2.7)
Score “E”: Latrine used by: Nobody in household uses latrine	4.5 (340/7489)	1.1 (13/1207)	-3.5 (-4.6, -2.3)*	1.0 (6/607)	-3.6 (-4.9, -2.3)	1.2 (7/600)	-3.4 (-4.7, -2.1)*
Score “8”: Latrine used by: No other household member – N/A	0.3 (21/7489)	0.9 (11/1207)	0.6 (0.1, 1.2)	0.7 (4/607)	0.4 (-0.2, 1.0)	1.2 (7/600)	0.9 (0.00, 1.8)
Assessing against narrative progress report: “Household members using”[§] AMONG ULTRA-POOR HHs	75.4 (1852/2457)	87.6 (359/410)	12.1 (7.2, 16.9)*	89.1 (180/202)	13.7 (7.8, 19.6)*	86.1 (179/208)	10.5 (4.3, 16.6)*

†RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VWC level; *Statistically significant at $\alpha=0.05$ ($p<0.05$)

§Our interpretation of “any household member using” included “latrine use, by whom” scores A-C amongst those with a QIS latrine coverage score of A, B, C, or D. We settled on this definition, as it most closely generated the proportions BRAC reported in their narrative progress report, and the wording of the report was vague.

Table 4a. QIS “latrine use, by whom” indicator HH04: comparison between BRAC’s QIS sample and the verification sample, and disaggregation between non-poor and ultra-poor/poor HHs

Indicators	BRAC’s QIS sample		Total verification sample		RD† between BRAC QIS and verification samples		Verification QIS sub-sample		RD† between BRAC QIS and verification QIS samples		Verification non-QIS sub-sample		RD† between BRAC QIS and verification non-QIS samples	
	% (n/N)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)	
	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs
QIS “latrine use, by whom” indicator HH04 (use amongst those who have a latrine [hygienic or unhygienic]):														
Score “A”: Latrine used by: Women and adolescent girls + children ≥6 years age + men and adolescent boys + feces of other household members end up in toilet	48.1 (1229/2555)	38.6 (1902/4934)	41.6 (173/416)	35.4 (280/791)	-6.3 (-13.2, 0.7)	-3.0 (-8.7, 2.7)	37.0 (77/208)	36.8 (147/399)	-10.8 (-20.1, -1.5)*	-1.5 (-8.0, 5.1)	46.2 (96/208)	33.9 (133/392)	-1.7 (-10.0, 6.6)	-4.5 (-12.1, 3.0)
Score “B”: Latrine used by: Women and adolescent girls + children ≥6 years age + men and adolescent boys	37.2 (951/2555)	37.1 (1830/4934)	45.9 (191/416)	48.9 (387/791)	8.4 (1.4, 15.3)*	11.3 (5.1, 17.5)*	50.0 (104/208)	46.1 (184/399)	12.4 (2.8, 22.1)*	8.3 (0.9, 15.8)*	41.8 (87/208)	51.8 (203/392)	4.3 (-3.9, 12.5)	14.3 (6.0, 22.5)
Score “C”: Latrine used by: Women and adolescent girls + children ≥6 years age	6.2 (159/2555)	9.1 (451/4934)	3.9 (16/416)	5.3 (42/791)	-2.3 (-4.7, 0.0)	-3.6 (-6.2, -1.0)*	3.9 (8/208)	7.3 (29/399)	-2.3 (-5.3, 0.6)	-1.6 (-5.0, 1.8)	3.9 (8/208)	3.3 (13/392)	-2.3 (-5.3, 0.6)	-5.6 (-8.5, -2.6)*
Score “D”: Latrine used by: Women and adolescent girls	6.1 (156/2555)	9.1 (450/4934)	8.4 (35/416)	7.5 (59/791)	2.3 (-1.6, 6.3)	-1.5 (-4.3, 1.3)	9.1 (19/208)	7.3 (29/399)	3.1 (-2.4, 8.6)	-0.7 (-5.4, 2.0)	7.7 (16/208)	7.7 (30/392)	1.6 (-3.5, 6.7)	-1.3 (-4.5, 2.0)
Score “E”: Latrine used by: Nobody in household uses latrine	2.2 (55/2555)	5.8 (285/4934)	0.2 (1/416)	1.5 (12/791)	-1.9 (-2.9, -1.0)*	-4.3 (-5.8, -2.7)*	0.0 (0/208)	1.5 (6/399)	-	-4.2 (-6.0, -2.4)	0.5 (1/208)	1.5 (6/392)	-1.7 (-2.9, -0.5)*	-4.3 (-6.0, -2.6)*
Score “8”: Latrine used by: No other household member	0.2 (5/2555)	0.3 (16/4934)	0.0 (0/416)	1.4 (11/791)	-	1.1 (0.2, 1.9)*	0.0 (0/208)	1.0 (4/399)	-	0.7 (-0.2, 1.6)	0.0 (0/208)	1.8 (7/392)	-	1.5 (0.1, 2.8)*
Assessing against narrative progress report: “Household members using”[§]	87.0 (2223/2555)	78.3 (3863/4934)	90.9 (378/416)	88.2 (698/791)	3.8 (-0.5, 8.2)	9.9 (6.0, 13.8)*	90.9 (189/208)	88.7 (354/399)	3.8 (-2.0, 9.7)	10.3 (5.3, 15.4)*	90.9 (189/208)	87.8 (344/392)	3.8 (-1.5, 9.2)	9.5 (5.2, 13.9)*

†RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VWC level; *Statistically significant at $\alpha=0.05$ ($p<0.05$)

§Our interpretation of “any household member using” included “latrine use, by whom” scores A-C amongst those with a QIS latrine coverage score of A, B, C, or D. We settled on this definition, as it most closely generated the proportions BRAC reported in their narrative progress report, and the wording of the report was vague.

QIS “latrine use, when” indicator (i.e., when [during which times of year/day] people use the HH latrine)

As shown in Table 5, significant differences in QIS score distributions along the “latrine use, when” ladder were observed between BRAC’s sample and the verification sample. The distinction between ladder scores “A” and “B” for this QIS indicator are quite small, with latrine utilization during “abnormal situations” as the defining factor between the two scores. It is not surprising, therefore, that the largest differences between BRAC’s QIS ladder scores and the verification ladder scores were observed in scores “A” and “B”. As with the latrine coverage and “latrine use, by whom” QIS indicators, internal consistency was observed in the verification sample for the “latrine use, when” QIS indicator in that similar deviations from BRAC’s QIS ladder score distributions were observed in the total verification and the verification QIS/non-QIS sub-groups.

In assessing the reported “latrine use, when” benchmark amongst ultra-poor households, as indicated by BRAC in their narrative progress report, a significantly higher proportion of ultra-poor verification households scored at or above benchmark compared to BRAC’s ultra-poor QIS households (95.6 vs. 80.7%, RD=14.6 [10.9, 18.3]). For comparisons between BRAC’s QIS sample and the verification QIS sub-group, and BRAC’s QIS sample and the verification non-QIS sample, the verification sub-groups had a higher proportion of ultra-poor households at or above benchmark compared to BRAC’s QIS sample for “latrine use, when” (95.5 vs. 80.7%, respectively; RD=14.6 [10.7, 18.4] and 95.7 vs. 80.7%, respectively; RD=14.6 [10.1, 19.1]). Readers should note that data presented in the first five rows of Table 5 represent a direct comparison between the latrine “use, when” ladder scores from BRAC’s QIS dataset against our verification dataset, while the data presented in the last row represents a verification of BRAC’s reported results, as per the wording in their narrative progress report.

Table 5a presents results with non-poor and ultra-poor/poor disaggregations for each level of the latrine “use, when” ladder scale. This table indicates some effect modification between non-poor vs. poor and ultra-poor households.

Table 5. QIS “latrine use, when” indicator (HH05): comparison between BRAC’s QIS sample and the verification sample, and BRAC QIS sample and verification QIS and non-QIS sub-samples

QIS indicators	BRAC’s QIS sample	Total verification sample	RD [†] between BRAC QIS and verification samples	Verification QIS sub-sample	RD [†] between BRAC QIS and verification QIS samples	Verification non-QIS sub-sample	RD [†] between BRAC QIS and verification non-QIS samples
	% (n/N)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)	% (n/N)	RD (95% CI)
QIS “latrine use, when” indicator HH05 (use amongst those who have a latrine [hygienic or unhygienic]):							
Score “A”: Latrine used: During the day and night of dry season + during the day and night of rainy season + during abnormal situations	58.9 (4413/7489)	89.7 (1083/1207)	30.9 (25.5, 36.3)*	88.6 (538/607)	29.7 (23.5, 35.9)*	90.8 (545/600)	32.1 (26.0, 38.2)*
Score “B”: Latrine used: During the day and night of dry season + during the day and night of rainy season	23.8 (1779/7489)	7.3 (88/1207)	-16.4 (-20.9, -12.0)*	8.1 (49/607)	-15.6 (-21.0, -10.3)*	6.5 (39/600)	-17.2 (-22.2, -12.3)*
Score “C”: Latrine used: During the day and night of dry season	9.4 (707/7489)	1.3 (16/1207)	-8.2 (-10.2, -6.2)*	1.7 (10/607)	-7.9 (-10.0, -5.7)*	1.0 (6/600)	-8.5 (-10.6, -6.4)*
Score “D”: Latrine used: During the day of dry season	3.6 (269/7489)	0.5 (6/1207)	-3.1 (-4.2, -2.0)*	0.5 (3/607)	-3.1 (-4.2, -1.9)*	0.5 (3/600)	-3.1 (-4.3, -1.8)*
Score “E”: Open defecation, latrine not used	4.3 (321/7489)	1.2 (14/1207)	-3.2 (-4.4, -1.9)*	1.2 (7/607)	-3.1 (-4.5, -1.8)*	1.2 (7/600)	-3.2 (-4.6, -1.8)*
Assessing against narrative progress report: “Regular use of the latrine”[§] AMONGST ULTRA-POOR HHs	80.7 (1982/2457)	95.6 (392/410)	14.6 (10.9, 18.3)*	95.5 (193/202)	14.6 (10.7, 18.4)*	95.7 (199/208)	14.6 (10.1, 19.1)*

[†]RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VWC level; * Statistically significant at $\alpha=0.05$ ($p<0.05$)

[§]Our interpretation of “regular use of latrine” included “latrine use, when” scores A-C amongst those with a QIS latrine coverage score of A, B, C, or D. We settled on this definition, as it most closely generated the proportions BRAC reported in their narrative progress report, and the wording of the report was vague.

Table 5a. QIS “latrine use, when” indicator HH05: comparison between BRAC’s QIS sample and the verification sample, and disaggregation between non-poor and ultra-poor/poor HHs

Indicators	BRAC’s QIS sample		Total verification sample		RD† between BRAC QIS and verification samples		Verification QIS sub-sample		RD† between BRAC QIS and verification QIS samples		Verification non-QIS sub-sample		RD† between BRAC QIS and verification non-QIS samples	
	% (n/N)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)		% (n/N)		RD (95% CI)	
	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs	Non-poor HHs	Ultra-poor & poor HHs
QIS “latrine use, when” indicator HH05 (use amongst those who have a latrine [hygienic or unhygienic]):														
Score “A”: Latrine used: During the day and night of dry season + during the day and night of rainy season + during abnormal situations	64.9 (1658/2555)	55.8 (2755/4934)	92.8 (386/416)	88.1 (697/791)	27.7 (22.0, 33.4)*	32.3 (26.4, 38.2)*	92.3 (192/208)	86.7 (346/399)	27.2 (19.8, 34.6)*	30.8 (24.0, 37.6)*	93.3 (194/208)	89.5 (351/392)	28.1 (22.4, 33.9)*	33.7 (26.9, 40.6)*
Score “B”: Latrine used: During the day and night of dry season + during the day and night of rainy season	24.2 (619/2555)	23.5 (1160/4934)	6.5 (27/416)	7.7 (61/791)	-17.6 (-22.6, -12.7)*	-15.8 (-20.5, -11.1)*	6.7 (14/208)	8.8 (35/399)	-17.4 (-23.9, -10.8)*	-14.8 (-20.5, -9.2)*	6.3 (13/208)	6.6 (26/392)	-17.9 (-23.1, -12.7)*	-16.8 (-22.3, -11.2)*
Score “C”: Latrine used: During the day and night of dry season	6.3 (161/2555)	11.1 (546/4934)	0.5 (2/416)	1.8 (14/791)	-5.7 (-7.5, -4.0)*	-9.3 (-11.7, -7.0)*	1.0 (2/208)	2.0 (8/399)	-5.3 (-7.3, -3.2)*	-9.0 (-11.6, -6.4)*	0.0 (0/208)	1.5 (6/392)	-	-9.6 (-12.1, -7.1)*
Score “D”: Latrine used: During the day of dry season	2.2 (55/2555)	4.3 (214/4934)	0.0 (0/416)	0.8 (6/791)	-	-3.5 (-4.9, -2.1)	0.0 (0/208)	0.8 (3/399)	-	-3.5 (-5.0, -2.1)*	0.0 (0/208)	0.8 (3/392)	-	-3.5 (-5.2, -1.9)*
Score “E”: Open defecation, latrine not used	2.4 (62/2555)	5.3 (259/4934)	0.2 (1/416)	1.6 (13/791)	-2.2 (-3.2, -1.2)*	-3.6 (-5.2, -2.0)*	0.0 (0/208)	1.8 (7/399)	-	-3.4 (-5.3, -1.6)*	0.5 (1/208)	1.5 (6/392)	-2.0 (-3.3, -0.7)*	-3.8 (-5.7, -1.8)*
Assessing against narrative progress report: “Regular use of the latrine” §	90.0 (2299/2555)	82.6 (4073/4934)	99.3 (413/416)	96.2 (761/791)	9.3 (7.0, 11.6)*	13.6 (10.6, 16.6)*	100.0 (208/208)	95.7 (382/399)	-	13.0 (9.6, 16.4)*	98.6 (205/208)	96.7 (379/392)	8.6 (5.9, 11.2)*	14.2 (10.9, 17.5)*

†RD: Cluster-adjusted risk difference generated using GEE - standard error adjusted for clustering on VVC level; *Statistically significant at $\alpha=0.05$ ($p<0.05$)

§Our interpretation of “regular use of latrine” included “latrine use, when” scores A-C amongst those with a QIS latrine coverage score of A, B, C, or D. We settled on this definition, as it most closely generated the proportions BRAC reported in their narrative progress report, and the wording of the report was vague.

3.1.2.3 Self-reported latrine utilization captured via household use schedule

Verification households reported an average of 6.2 household-level latrine users (IQR 4, 7), with a range of 1 to 28 latrine users per household. The number of actual household members in these households averages 4.7 persons (IQR 4, 6), but ranges from 1 to 24, with other regular latrine users (e.g., neighbors, tenants, servants) ranging from 0 to 4 persons per surveyed household. The number of total daily latrine events self-reported by verification households averages 8.3 (IQR 5, 10), but ranges from 0 to 47. After adjusting for sampling design to get population estimates, we find that households reported the highest level of latrine utilization during the morning hours, followed by evening, afternoon, and nighttime. See Table 6 for further details.

Table 6. Daily latrine utilization self-reported by verification households

Time of day [†]	Average HH-level self-reported use	Range of HH-level self-reported use	Average individual-level* self-reported use	Range of individual-level* self-reported use
	n (95% CI)	(min, max)	n (95% CI)	(min, max)
Morning	6.02 (5.75, 6.28)	(0, 28)	0.97 (0.96, 0.98)	(0, 4)
Afternoon	0.58 (0.50, 0.66)	(0, 8)	0.09 (0.08, 0.11)	(0, 2)
Evening	1.22 (1.03, 1.40)	(0, 19)	0.19 (0.16, 0.22)	(0, 2)
Night	0.49 (0.41, 0.58)	(0, 14)	0.09 (0.07, 0.10)	(0, 4)
Daily Total	8.3 (7.90, 8.72)	(0, 47)	5.38 (5.20, 5.55)	(0, 19.2)‡

* Use for any given household latrine user

† Morning defined 04:00-10:00, afternoon defined 10:00-15:00, evening defined 15:00-19:00, night defined 19:00-04:00

‡ Average daily range per household latrine users

In preparation for our analysis of the comparison between self-reported latrine events and instrument-recorded defecation events, we multiplied this daily number of latrine events by four to produce a four-day household-level total for self-reported latrine events. Only the four-day household total self-reported latrine utilization data from the sub-set of PLUM households were used to compare self-reported utilization to PLUM-recorded defecation. The four-day total number of self-reported latrine events from this sub-set averages 32.8 (28.6, 37.0) events, and ranges from 8 to 140 events. Figures 9 and 10 present distributions of daily household-level self-reported latrine events amongst all verification households, and four-day total household-level self-reported latrine events amongst the sub-set of PLUM households.

Figure 9. Distribution of household-level self-reported daily latrine events for all verification HHs

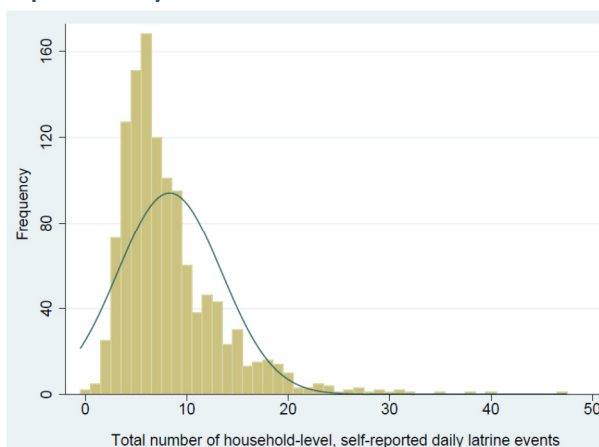
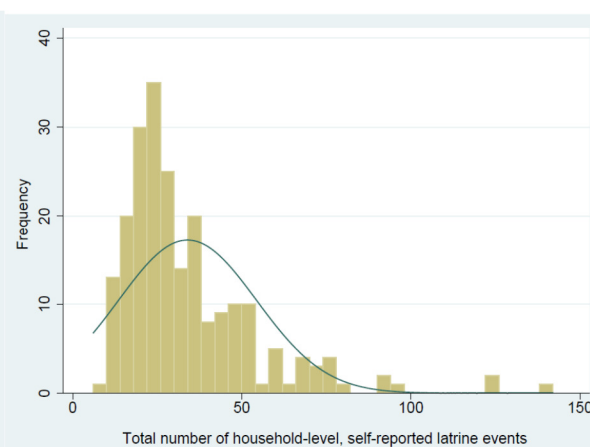


Figure 10. Distribution of four-day, household-level self-reported latrine events for the sub-set of PLUM HHs



Household-level utilization designations

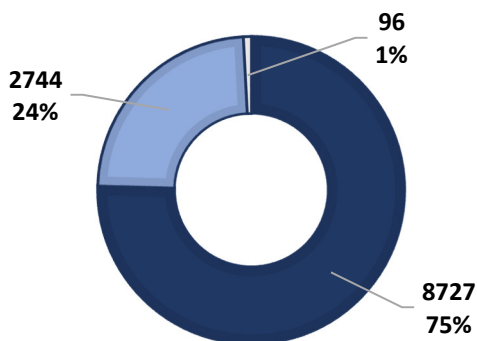
Amongst the verification sample, 75.6% (913/1207) of households report that all members always exclusively use a latrine for defecation. We term these households as “always using” households. All 5147 latrine users from these 913 households reportedly always exclusively defecating in a latrine. However, 22 of the 5147 latrine users from the 913 “always using” households reported their primary place of defecation as only “most of the time in the latrine” as opposed to “always in latrine”. We left these households in the “always using” category based on our pre-determined decision to make the “exclusive defecation in a latrine” variable the decisive variable for household utilization designation. Another two households reported all latrine users as “always exclusively” defecating in the latrine, but one latrine user in each household reported open defecation as their primary place for defecation. As a result of the nature of these discrepancies, we moved these households to the “sometimes using” category. Finally, another 12 households (i.e., 927 total households) report that all members always use the latrine as the primary place for defecation, but when asked if they always exclusively used that latrine for defecation, at least one or more household latrine users indicated that they did not always exclusively use the latrine for defecation. These households did not meet the definition of “always using”, as all latrine users did not reportedly always exclusively use the latrine for defecation. On weighted analysis of the verification sample, the “always using” household latrine utilization category is comprised of 27.1% (2365/8727) ultra-poor households, 28.6% (2499/8727) poor households, and 44.3% (3863/8727) non-poor households.

Households report some level of latrine use for defecation amongst household latrine users in 23.5% (284/1207) of verification households. We term these households as “sometimes using” households. Of the 2324 latrine users from these 284 “sometimes using” households, 80.0% (1860/2324) reportedly always exclusively defecate in a latrine, 12.2% (284/2324) never use a latrine, 4.4% (103/2324) only sometimes defecate in a latrine, 1.8% (41/2324) sometimes openly defecate (46.3% [19/41] sometimes openly defecate outside the home compound, and 53.7% [22/41] sometimes openly defecate within the compound), and respondents reported not knowing the defecation habits of 1.5% (36/2324) of latrine users in these households. On weighted analysis of the verification sample, the “sometimes using” household latrine utilization category is comprised of 24.3% (668/2744) ultra-poor households, 29.7% (816/2744) poor households, and 45.9% (1261/2744) non-poor households.

Amongst ultra-poor households with access to a latrine scoring A-D on the QIS sanitation ladder (those having a hygienic or unhygienic latrine), 76.9% (303/394) were classified as “always using” households, and 100.0% (394/394) were classified either “always using” or “sometimes using” a latrine for defecation, per our pre-determined household-level utilization designations (i.e., 17 ultra-poor households scored “E” on the ladder scale). While it is possible to make a loose comparison between these self-reported use data and the QIS “latrine use, by whom” and “latrine use, when” data presented in the last row of Tables 4 and 5 (87.6% [359/410] of ultra-poor households with access to latrines in our verification sample had household members using the latrine, per the QIS ladder scoring system; 95.6% [392/410] of ultra-poor households with access to latrines in our verification sample had household members “regularly using” the latrine, per the QIS ladder scaling system), caution should be taken when interpreting these findings. The data are not directly comparable, as the QIS latrine coverage ladder score “E” encompasses both households that do not have a latrine and households that have a latrine without a ring or a slab. Therefore, when we try to compare household-level use designation derived from the employment of a household use schedule with the QIS ladder scores, we will lose data on those households that do have a latrine, but still scored an “E” for their latrine coverage indicator (i.e., these households are not captured in the denominator, therefore proportions are higher than they should be due to the definition of score “E”). In addition, the self-reported use data captured via our household use schedule represents a systematic combination of data on the demographics of latrine users as well as their frequency of use

information captured by both the “latrine use, by whom” and “latrine use, when” indicators (i.e., the household use schedule not only captures information about who is using the latrine, but also when they use the latrine). See section 4.3 for further discussion on this point and additional comparisons of these data with PLUM-recorded use data.

Figure 11. Population-averaged household-level utilization designations



■ Always using HHs ■ Sometimes using HHs ■ Never using HHs

Households report all members never exclusively use a latrine for defecation in 0.8% (10/1207) of verification households. We term these households as “never using” households. It is worth noting that only one household that reported all latrine members as “never exclusively using” a latrine actually has access to a household latrine. Amongst the 40 latrine users from these ten households, 77.5% (31/40) reportedly always openly defecate outside the household compound, 17.5% (7/40) reportedly open defecate outside of the household compound most of the time, and 5% reportedly always openly defecate within the household compound. On weighted analysis of the verification sample, the “never using” household latrine utilization category is comprised of 47.4% (45/95) ultra-poor households, 32.4% (31/95) poor households, and 20.2% (19/95) non-poor households. Figure 11 graphically presents population-averaged estimates of household-level utilization designations in areas surveyed during the verification.

3.1.2.4 Instrument-recorded latrine utilization captured via PLUMs

The team successfully installed PLUMs in 250 of 319 households targeted for PLUM installation, resulting in a PLUM installation success rate of 78%. Amongst households in which a PLUM was successfully installed, two households tampered with the sensor installed in their latrine, and opened it prior to the retrieval date, resulting in a loss of data. In addition to data loss due to households tampering with the PLUMs, we also lost data from 31 latrine installations from 21 households as a result of technical issues that prevented proper electronic data capture. Therefore, our final analytical PLUM dataset consisted of instrument-recorded signal data from a total of 231 latrine installations at 220 households. Figure 12 outlines the flow for PLUM data capture.

Cluster breakdowns of installation refusal

PLUM installation was unsuccessful in 69 households:¹¹

- A total of 43 households (62%, 43/69¹² of households unsuccessful for PLUM installation) had a latrine with a superstructure that did not support PLUM installation;
- A total of 22 households (32%, 22/69¹³) completely refused PLUM installation;¹⁴
 - The primary reason for refusal was religious issues (86.4% [19/22]), followed by a belief the PLUM was a camera (9.1% [2/22]), and other privacy-related concerns (4.5% [1/22]).
- Three households (4%, 3/69) did not have access to a latrine, and practiced open defecation; and

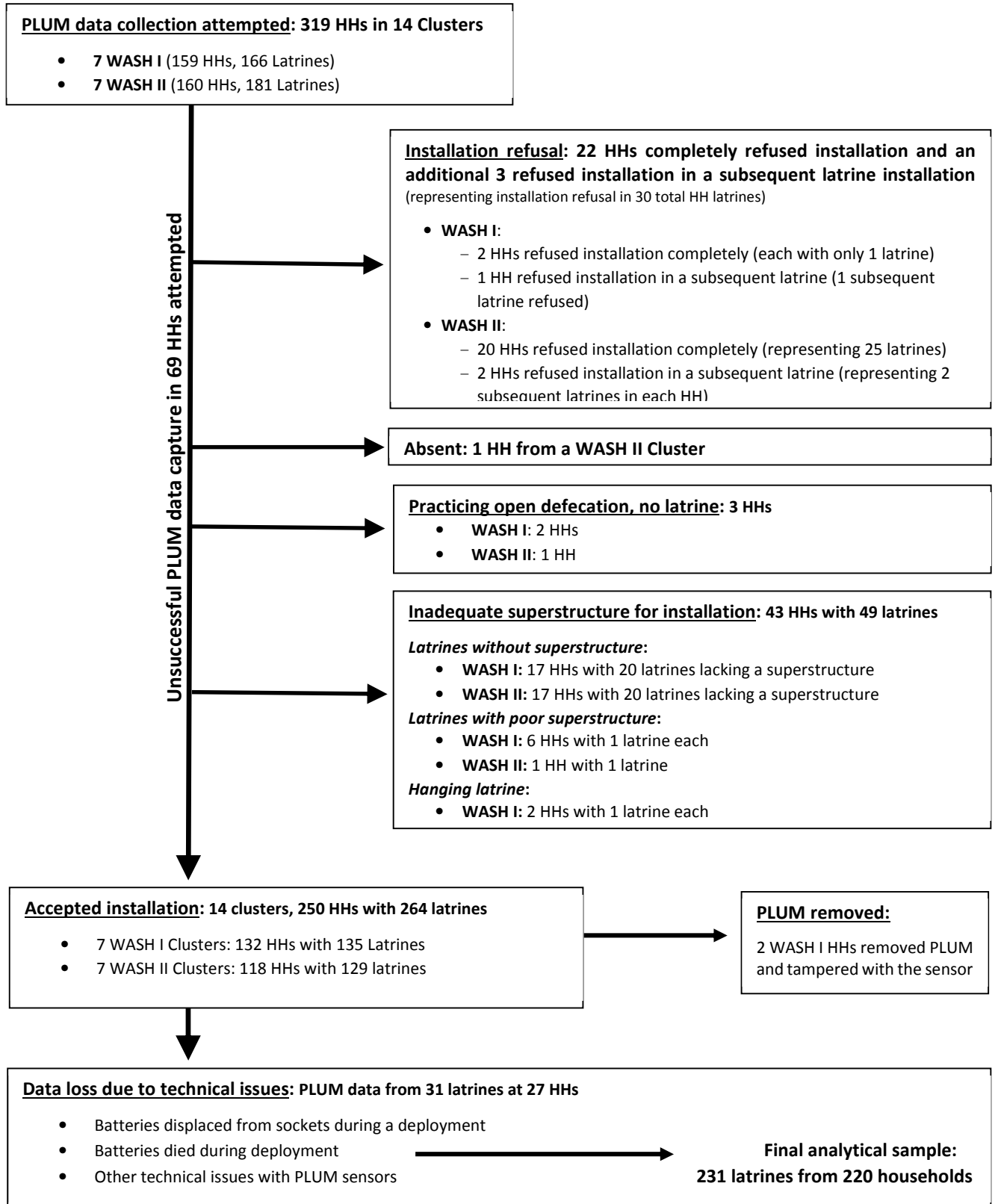
¹¹ The 69 households had a total of 80 latrines (i.e., some households had multiple latrines, and our protocol indicated installation of a PLUM in every functional latrine in selected clusters).

¹² Each household had one latrine.

¹³ The 22 households had a total of 27 latrines.

¹⁴ An additional three households accepted PLUM installation in one latrine, but refused installation in a subsequent latrine.

Figure 12. PLUM data capture



- One household (1%, 1/69) was absent.

One or more households refused PLUM installation in five clusters randomly targeted for PLUM installation. Amongst those five clusters, a total of 22 households completely refused PLUM installation for all latrines (termed herein as “complete [PLUM] refusal”), representing a sub-sample refusal rate of 6.9% (22/319 households) amongst households targeted for PLUM data collection. An additional three households accepted PLUM installation in one latrine, but refused PLUM installation in a subsequent latrine (termed herein as [PLUM] “refusal”), representing a refusal rate for any installation of 7.8% (25/319 households). Cluster and intra-cluster refusal varied amongst the five clusters:

- The highest PLUM installation refusal rate was observed in Ahmed H. Bari VWC, where 17 households refused PLUM installation in at least one latrine, 15 of which completely refused PLUM installation. This represents a cluster refusal rate of 71% (17/24) for refusal of installation in at least one latrine, and a 63% (15/24) refusal rate for complete refusal. All refusing households cited religious concerns as the reason for refusal in this cluster.
- The next highest PLUM installation refusal rate was observed in Azadi Bazar VWC, where four households¹⁵ completely refused PLUM installation. The cluster refusal rate there was 17% (4/24).
- Two households completely refused PLUM installation in one cluster (for an intra-cluster refusal rate of 8% [2/24]), and one household completely refused PLUM installation in two clusters.

PLUM-recorded defecation events

After accounting for data loss due to the variety of reasons outlined above, PLUM data were captured on 217 primary (i.e., most frequently used household latrine, per household respondent) latrines, and 14 secondary (i.e., second most frequently used household latrine) latrines. In total, during the four-day analytical period, the number of defecation events captured in all household latrines in which PLUMs were installed ranged from 0 to 107, with an average of 21.7 (95% CI 18.1, 25.4) defecation events detected per household. Figure 13 graphically displays the distribution of four-day, household-level PLUM-recorded defecation events amongst the sub-set of PLUM households. Given there is an average of 6.2 (5.6, 6.8) latrine users per household in this sub-set, this seems within the expected range of household totals assuming each member defecated in the latrine once per day.

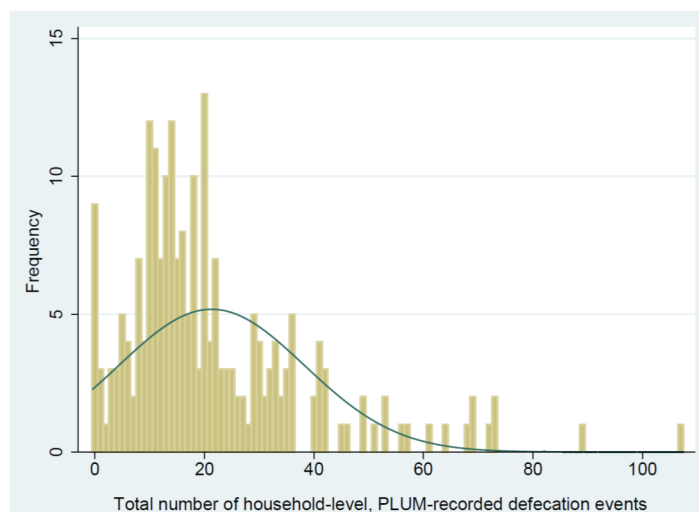
The number of defecation events captured in primary latrines per day ranged from 0 to 28 over the four-day PLUM analytical period, with a daily average number of defecation events per latrine ranging from 5.1 (95% CI 4.1, 6.1) to 5.5 (4.5, 6.4). Instrument-recorded defecation data were relatively stable over the four-day analytical period, as there was no particular trend in the number of defecation events detected over time. The number of defecation events captured in secondary latrines per day ranged from 0 to 16 over the four-day PLUM analytical period, with a daily average number of defecation events per latrine ranging from 1.9 (95% CI 0.6, 3.1) to 2.8 (1.1, 4.5). As with primary latrines, instrument-recorded defecation data captured in secondary latrines were relatively stable over the four-day analytical period, as there was no particular trend in the number of defecation events detected over the four days.

In order to determine whether there was a difference in sanitation outcomes between households with PLUM failures and households included in our analytical PLUM sample, we compared the number of household-level self-reported latrine events obtained via household use schedule across these two sub-groups. The number of household-level latrine use events did not differ significantly between these two groups (RR=1.1 [95% CI 0.8, 1.6, p=0.585]). On average, households with PLUM failures reported 33.9

¹⁵ The four households had a total of five latrines.

latrine events (IQR 20, 40), and households included in the analytical PLUM sample (i.e., those with no PLUM failures) reported 37.6 latrine events (IQR 20, 52).

Figure 13. Distribution of four-day, household-level PLUM-recorded defecation events



3.1.2.5 Latrine spot check indicators captured via direct observation

Presence of latrine spot check indicators are summarized in Table 7. A well-worn path to the latrine, a wet latrine floor, presence of fecal odor, discoloration of the pan or slab, and visible feces were amongst the most commonly observed latrine spot check indicators. A more detailed discussion regarding the presence of latrine spot check indicators and observed use can be found in sections 3.5.2 and 4.4.

Table 7. Latrine cleanliness and utilization spot check indicators

Latrine spot check indicator	Households (un-weighted)	Population-averaged (weighted)	
	Latrine spot check outcome	Latrine spot check outcome	95% CI
	n (%)	n (%)	
Latrine cleanliness			
Stagnant water visible over latrine slab and/or floor	138 (11.6)	1,479.5 (12.9)	1113.2, 1845.8
Visible feces in latrine pan, or on the slab or floor observed	662 (55.4)	6,126.0 (53.5)	5097.2, 7154.7
Visible discoloration of pan or slab observed	692 (58.0)	6,702.4 (58.5)	5502.5, 7902.2
Smell of feces observed	752 (63.0)	7,002.0 (61.2)	5858.0, 8146.0
Dirt, leaves, spider webs observed in latrine pan/pit	331 (27.7)	3,178.3 (27.8)	2344.4, 4012.2
Presence of flies observed in latrine	441 (36.9)	3,923.6 (34.3)	3174.9, 4672.3
Latrine cleaning agents observed	205 (17.2)	2,145.9 (18.7)	1581.6, 2710.2
Evidence of latrine use			
Well-worn path to the latrine observed	1,141 (95.6)	10,943.5 (95.6)	9563.7, 12323.3
Wet latrine floor (no excess stagnant water) observed	876 (73.4)	8,654.5 (75.6)	7422.0, 9886.9
Slippers for latrine use observed near latrine	242 (20.3)	2,417.6 (21.1)	1801.9, 3033.4
Available water for flushing or anal cleansing observed	380 (31.8)	3,892.8 (34.0)	3116.5, 4669.2
Available water near latrine for handwashing observed	441 (36.9)	4,515.8 (39.4)	3714.6, 5317.0
Handwashing agents (e.g., soap, soapy water, ash) observed	396 (33.2)	3,967.9 (34.7)	3176.9, 4758.8
Evidence latrine was used for purposes not related to sanitation	3 (0.3)	50 (0.4)	-23.3, 123.3
Number	1194	11450	--

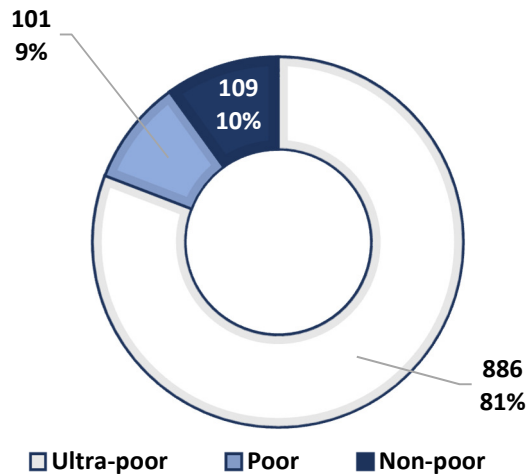
Notes: CI: confidence interval

3.2 VERIFICATION QUESTION 2: Is BRAC achieving its poverty targeting goals (i.e., are ultra-poor households receiving grants/subsidies for latrine construction, are poor and ultra-poor households receiving loans for latrine construction and repair?)

3.2.1 Intervention receipt by wealth category: self-reported receipt of outside assistance

3.2.1.1 Self-reported receipt of outside assistance for latrine construction

Figure 14. Population-averaged estimates for receipt of BRAC latrine construction support amongst households with an improved or shared but otherwise improved latrine



Amongst households with an improved or shared but otherwise improved latrine reportedly constructed since 2011, 45.4% (171/377) reported receiving outside assistance for the latrine construction. Of these 171 households, 134 (78.4%) reported receiving assistance from BRAC for construction of one

or more latrines since the beginning of BRAC's WASH project in 2011. Wealth status is strongly associated with receipt of latrine construction assistance from BRAC during the WASH II project period ($p < 0.001$), with poor and non-poor households with a lower probability of receiving assistance from BRAC (crude RR=0.24 [0.15, 0.41] and crude RR=0.18 [0.10, 0.34], respectively). Figure 14 presents population-averaged estimates for latrine construction assistance from BRAC for the construction of an improved or improved but otherwise shared latrine during the project period for areas surveyed during the verification. These figures indicate that the majority (81%) of the households in surveyed areas that have an improved or shared but otherwise improved latrine (i.e., the type of latrines BRAC promotes under the WASH II program) and reported receiving latrine construction support from BRAC are ultra-poor households. While BRAC aims to target only ultra-poor households with direct latrine construction support, a small proportion of non-poor and poor households with an improved or improved but otherwise shared latrine reported receiving latrine construction support from BRAC (10% and 9% for non-poor and poor households, respectively). These reports may indicate a misdirection of program interventions, a misclassification of household wealth status at the VWC level, or respondent recall bias.

Five households reporting receipt of support from BRAC for the construction of their improved or shared but otherwise improved latrine since 2011 reported receiving loan support for the construction of the latrine – two of which are ultra-poor households, two of which are poor households, and one of which is a non-poor.

3.2.1.2 Self-reported receipt of outside assistance for latrine repair

Households with improved or shared but otherwise improved latrines repaired during the WASH project period represent 13.9% (168/1207) of households in the verification sample. The 168 verification

households that repaired an improved or shared but otherwise improved latrine during BRAC's WASH project period reported repairing a total of 170 latrines. The 46 ultra-poor households reported a total of 46 improved or shared but otherwise improved household latrines were repaired during the project period, of which all 46 households reported the repair of only one household latrine. Of the ultra-poor households reporting the repair of an improved or shared but otherwise improved latrine during the WASH project period, 97.8% (45/46) reported the repair of the most frequently used household latrine. The 57 poor households reported a total of 57 improved or shared but otherwise improved household latrines were repaired during the project period, of which 57 households reported the repair of one latrine. Of the poor households reporting the repair of an improved or shared but otherwise improved latrine during the WASH project period, 94.7% (54/57) reported the repair of the most frequently used household latrine. The 65 non-poor households reported a total of 66 improved or shared but otherwise improved household latrines were repaired during the project period. Of the non-poor households reporting the repair of an improved or shared but otherwise improved latrine during the WASH project period, 89.2% (58/65) reported the repair of the most frequently used household latrine.

While 16 surveyed households reported receiving outside assistance for latrine repairs since 2011, only 11 households (6.5% [11/168]) with an improved or shared but otherwise improved latrine repaired since 2011 reported receiving outside assistance for the repairs. No households that reported receiving outside support for latrine repairs reported receiving loan support from any entity for the repairs.

3.3 VERIFICATION QUESTION 3: Are latrine coverage and utilization similar amongst project participants receiving support via grants/subsidies compared to those receiving loans?

Only six households surveyed reported loan receipt for the construction of their improved or shared but otherwise improved latrine since 2011, and only five of those households reported loan receipt from BRAC. Due to the small number of households reporting receipt of loans for latrine construction and/or repair, the sample is too small to be powered to detect differences in latrine coverage and utilization.

3.4 VERIFICATION QUESTION 4: How do various measures of latrine utilization (i.e., QIS ladder scores, self-reported use captured via structured household use schedules, instrument-recorded use, and latrine spot check indicators assessed through direct observation) compare?

3.4.1 Comparisons of latrine utilization measurement methods

3.4.1.1 Comparison of self-reported and instrument-recorded utilization

As indicated in Figure 15, there is a positive correlation between observed PLUM-recorded and self-reported latrine utilization. The four-day, household-level total number of self-reported latrine events is, however, significantly different than the four-day total of instrument-recorded defecation events ($p < 0.001$). After adjusting for survey design, households that had a PLUM installed secondary to survey administration were found to self-report a four-day average of 32.8 events (95% CI 28.6, 37.0) vs. a four-day average of 21.7 events (95% CI 18.1, 25.4) recorded with the PLUMs. This suggests over-reporting of self-reported latrine utilization.

In order to account for the clustering in our observed survey data while making comparisons between measurement methods, we generated a simplistic GEE model comparing the observed household-level total number of defecation events detected by the PLUM during the four-day analytical period (our

independent variable), and the observed four-day, household-level total number of self-reported latrine events (our dependent variable). We then used this crude model to estimate the expected number of PLUM-recorded defecation events and self-reported latrine events from the observed values in the sample sub-set, and generate a fitted probability line and accompanying 95% confidence interval. Figure 16 presents the outcome of model-estimated expected values. Households in our verification sample reported a considerable amount of latrine utilization that was not detected by the PLUM as a defecation event, as indicated by an average intercept of 28.1 events, on crude analysis. This means that, on average, household respondents reported an average of 28.1 latrine events when zero PLUM-recorded defecation events were detected. The relative change in the average number of events in this comparative analysis can help quantify the exaggeration of self-reported latrine use. While a fitted line was produced from this simplistic GEE model, the model is not fully adjusted, and therefore may not appropriately represent the fully adjusted relative change in the average number of self-reported latrine events per PLUM-recorded defecation event. Additional investigation is needed in order to determine the exaggeration in self-reported utilization from a fully adjusted model.

Figure 15. Observed four-day total self-reported latrine and PLUM-recorded defecation events amongst the sub-set of households with PLUM installation

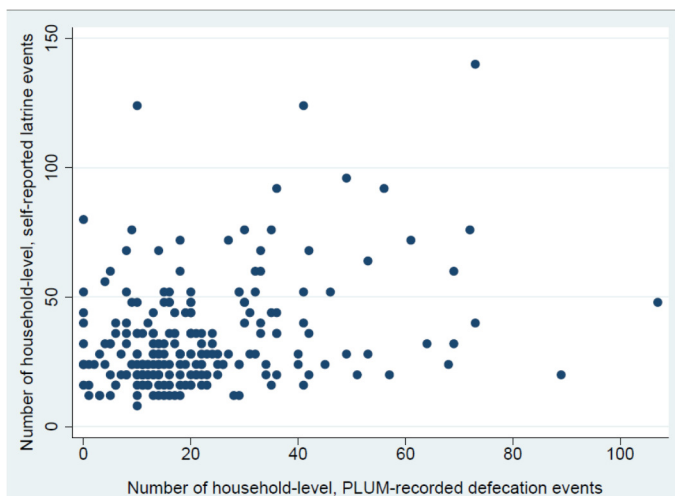
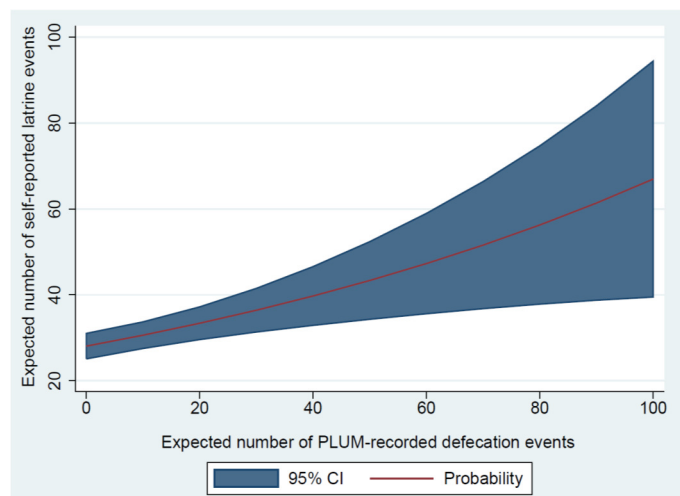


Figure 16. Expected four-day total self-reported latrine vs. PLUM-recorded defecation events amongst the sub-set of households with PLUM installation



3.4.1.2 Association between latrine spot check indicators and QIS latrine utilization indicators

Several latrine spot check indicators are associated with QIS “latrine use, by whom” and “latrine use, when” indicators. Table 8 summarizes those associations. Three latrine spot check indicators – visible discoloration of latrine pan or slab, available water for flushing or anal cleansing, and available water near latrine for handwashing – are associated with both QIS indicators. Readers should exercise caution in interpreting findings related to associations between latrine spot check and QIS latrine utilization indicators, as the sensitivity of the indicators needs to be taken into account when interpreting the data. See the discussion section for additional discussion on this point.

Table 8. Associations between latrine spot check indicators and other utilization measurement methods

	Association with QIS "latrine use, by whom" indicator ⁺	Association with QIS "latrine use, when" indicator ⁺
Latrine cleanliness		
Stagnant water visible over latrine slab and/or floor	p<0.001 *	p=0.732
Visible feces in latrine pan, or on the slab or floor observed	p=0.132	p<0.001 *
Visible discoloration of pan or slab observed	p=0.007 *	p<0.001 *
Smell of feces observed	p=0.122	p<0.001 *
Dirt, leaves, spider webs observed in latrine pan/pit	p=0.448	p<0.001 *
Presence of flies observed in latrine	p=0.008 *	p=0.080
Latrine cleaning agents observed	p=0.235	p=0.041*
Evidence of latrine use		
Wet latrine floor (no excess stagnant water or water log) observed	p=0.129	p<0.001 *
Slippers for latrine use observed near latrine	p=0.230	p=0.387
Available water for flushing or anal cleansing observed	p=0.022 *	p=0.015*
Available water near latrine for handwashing observed	p<0.001 *	p<0.001*
Presence of handwashing agents (e.g., soap, soapy water, ash)	p=0.211	p=0.035*
Notes: ⁺ Wald p-values, * statistically significant at $\alpha = 0.05$		

3.5 Cost assessment for various latrine utilization measurement methodologies

Our cost assessment revealed that instrument-recorded latrine utilization was the most expensive latrine utilization measurement method. These instrument-recorded data incur approximately five times the administration costs and nearly two and a half times the equipment rental costs of those incurred to obtain latrine spot check indicator data, and just over one and a half times of those incurred to obtain self-reported use data. While self-reported use data are cheaper to obtain than instrument-recorded use data, they incur three times the administration costs of those incurred to obtain latrine spot check indicator data. Although costs may influence decision-making related to latrine utilization measurement methodology, it is important to consider these costs in relation to the granularity of data and final products that are generated from the measurement method. Table 9 provides further details related to the cost assessment, and includes a section that provides an overview of data considerations. Additional discussion regarding these considerations can be found in section 4.6 *Latrine utilization measurement method cost considerations*.

Table 9. Summary of costs associated with various latrine utilization measurement methodologies

Measurement methodology	Monthly equipment rental	No. of units	Total person-hours	Average administration time per unit	Price per person-hour [†]	Crude cost per unit [‡]	Considerations
Instrument-recorded use measured via PLUM sensors	Price per PLUM: \$50/month [¥]	262 latrines	42	10 minutes per PLUM	\$2.96	\$0.49 / household	Data generalized to the household level; data may be influenced by possible reactivity and measurement biases
Self-reported use measured via household use schedule	Price per tablet: \$21/month [§]	1205 households	112	6 minutes per household	\$2.96	\$0.30 / household	Data allow for assessment of trends in individual and household level use; data may be influenced by possible administration, recall, and social desirability biases; evidence of over-reported self-reported use was observed on comparative analyses of the verification data
Latrine spot check indicators of use measured via visual inspection of latrines and surrounding premises	Price per tablet: \$21 month [§]	1193 households	38	2 minutes per household	\$2.96	\$0.10 / household	Data generalized to the household level; resulting binary outcome data are non-quantifiable [*] ; use indicators have varying levels of specificity; data influenced by possible administration bias

[†] Hourly rate for one fixed-term Field Research Assistant

[‡] Crude cost per unit only represents the costs associated with the administration of the methodology, exclusive of equipment rental costs

[¥] Monthly PLUM rental rate, inclusive of equipment rental and technical support

[§] Estimated monthly tablet rental rate, using the 2013 price of purchase, assuming a 2-year product life

^{*} Data are non-quantifiable in that they do not quantify latrine or defecation events

4. DISCUSSION

4.1 Comparability of results

With some exceptions, our survey results compare reasonably well with BRAC's reported results. This is to be expected, as we designed part of the verification to replicate the manner in which BRAC collects its QIS data (though with some changes to address potential issues with the sampling strategy). In general, the QIS ladder scores for latrine coverage and utilization generated from our verification sample were higher than those generated from BRAC's QIS sample. These increases in scores may be due to the amount of time that passed between BRAC's first round of QIS data collection and our verification data collection. During the year between data collection time points, BRAC continued implementing its WASH II project, during which progress was made to improve sanitation outcomes and household-level attributes (e.g., income, education, and migration) may have changed. In addition to the lapse of time between data collection points, seasonality may also have contributed to the differences in latrine coverage and utilization measures. There was less rain during the 2014 monsoon season compared to prior years, and the overall security situation was stable during our data collection period. Both the presence of rain and insecurity often results in decreased use of latrines, particularly during the night-time. These temporal changes may account for some of the increase in reported latrine utilization. Although decreases in metrics were rare, they were observed. These decreases may be explained by the destruction of hardware resulting from flooding, natural disasters, and poor household maintenance.

While our verification results compare reasonably well with BRAC's reported results, there is room for improvement of indicators and measurement methodologies. Anecdotally, we saw that the sanitation ladder scale definitions were not specific enough to capture the nuances of household utilization. During our field site visit, we observed differential interpretations of the ladder scale scores when during a meeting at the Kapasia regional office, we asked a question about how a particular household scenario (if some, but not all members of a cohort within the household were using the latrine) would be scored, different BRAC staff gave different responses. The QIS latrine utilization indicators are not well matched to produce the intended distributions, as indicated in the *QIS monitoring guidelines for the sample study 2012*, which indicates that ladder scales "A" and "B" are the "ideal, which possibly none or only a few households/schools/clusters can achieve". However, the mean QIS ladder score for both BRAC's QIS sample and the verification sample surpassed these ladder scores. If ladder scores "A" and "B" should truly indicate rare household scenarios, BRAC should consider modifying the QIS ladder score definitions.

4.2 Wealth category designations

Our results indicate a strong association between wealth status and reported receipt of latrine construction assistance from BRAC during the WASH II project period ($p < 0.001$), as 48.9% (886/1811) of ultra-poor, 6.2% (101/1625) of poor, and 6.2% (109/1765) of non-poor households with an improved or shared but otherwise improved latrine reported receiving latrine construction assistance from BRAC since 2011. We observed what was to be expected with regard to the distribution of household latrine construction amongst ultra-poor households, and the distribution of household latrine repairs amongst poor households since BRAC's WASH project targets ultra-poor households for subsidy receipt for latrine construction, and poor and ultra-poor households with loan receipt for construction and repairs. However, with 6.2% of poor and 6.2% of non-poor households (on weighted analysis, 21.6%, [29/134] – on un-weighted analysis) reportedly receiving subsidy support for latrine construction from BRAC, there is still some mis-targeting of project interventions.

Household wealth category designations were assigned via a BRAC-facilitated PRA activity and mapping exercise upon the VWCs' engagement in the WASH project (i.e., during 2007 for WASH I, and 2011-2012 for WASH II). The household wealth classifications assigned at the onset of each VWC's WASH I/WASH engagement are still being used by the project in these communities. According to BRAC staff, the project teams have not conducted a systematic update of wealth category classifications in two to seven years for WASH II and WASH I engaged communities, respectively.

During September 2011, VWCs updated their community maps and VWC registers with households that were new to the communities as of BRAC's 2007 census. These households were added to community maps and assessed for wealth status via PRA. During this process, however, BRAC and the community did not conduct a re-appraisal of the entire community (i.e., WASH I VWCs are still using 2007 wealth category classifications for most households, and 2011 wealth category classifications for the few households that were added to communities between 2007 and September 2011). BRAC facilitated a community PRA and mapping activity for VWCs engaged under WASH II during 2011-2012, but has not facilitated a systematic update of household wealth classification since the baseline PRA activity over two years ago.

4.3 Reported latrine use

Reported use was exceptionally high, with 75.6% (913/1207) of households reporting that all members always exclusively use a latrine for defecation. However, crude analysis indicated that, on average, 28.1 latrine events were reported when no PLUM defecation events were recorded. This raises questions about

reporting and courtesy biases. While caution should be taken when comparing QIS ladder scores against self-reported latrine use data obtained via household use schedule, as those data obtained via household use schedule result from a systematic combination of data on the demographics of latrine users as well as their frequency of use (i.e., a combination of the QIS “latrine use, by whom”, and “latrine use, when” indicators, these juxtapositions still provide an interesting comparison. When comparing these self-reported use data to the QIS latrine use indicators, self-reported use captured via household use schedule is found to be higher than QIS-assessed use captured via QIS ladder scales. This higher proportion of self-reported use via household use schedule may indicate capture of additional latrine use events that may not otherwise have been captured via the QIS ladder scale system, or the higher proportion may simply indicate systematic capture of over-reported use. Comparison of self-reported use data captured via household use schedule and PLUM-recorded use data suggests the higher proportion may likely be explained, at least to a certain extent, by the latter. Results from instrument-recorded use raises more serious questions about exaggerated use in self-reports. This is consistent with what instrumented monitoring has shown about water filters and cook-stoves (Thomas et al. 2014).

When interpreting instrument-recorded data, it is important to consider the pattern of PLUM installation approval amongst refusing households (see *section 4.5 Instrument-recorded use*). Refusing households, particularly ultra-poor households often accepted installation in the primary household latrine, but not secondary latrines. It is possible, therefore, that people shifted to using the secondary latrine (in which no PLUM was installed) during the PLUM data collection period. If this was the case, the difference between self-reported use and instrument-recorded use may not be completely accurate.

It is worth noting, however, that although the household use schedule included a series of questions regarding one’s defecation practices (e.g., primary place of defecation, whether the primary place of defecation changes over the course of the year, whether the person exclusively uses a latrine for defecation) the survey prompts that quantified latrine use at specific time periods throughout the day were translated to “how many times do [latrine users] use [the latrine]” during the specified periods. In other words, even though the three questions immediately preceding these prompts were specific to defecation, and the enumerator was instructed to read the survey directions to the respondent¹⁶, these questions did not specify latrine use for defecation purposes. Though this is a limitation of the survey instrument, in rural areas, latrines are typically limited to defecation (i.e., people tend to urinate in open places). The lack of specificity in these survey prompts may have contributed to the over-reporting of latrine utilization compared to PLUM-recorded utilization. An additional reporting bias may have been introduced in that we requested information on defecation practices for all persons using the household latrine, even if they were not present to provide responses for themselves. This may have contributed to over-reporting of latrine utilization.

4.4 Directly observed indicators of use

Indicators of use were largely consistent with reported use, but the sensitivity of some of the latrine use indicators should be considered when interpreting latrine spot check indicator results. For instance, the path to a household latrine may fall along the same pathway to other commonly visited places. While these indicators may provide an indication that the latrine is being used (by at least one person), the granularity of the data is not such that one can draw inferences to utilization patterns amongst different individuals or age-groups within the household. The indicators may not do much in the way of providing

¹⁶ The instructions for this section of the survey were as follows: Please provide the usual place of defecation (throughout the year) of all members of your household, starting with the eldest, and moving to the next eldest, and so forth. Please include information on feces disposal practices for all children.

information regarding the frequency and consistency with which the latrines are being used (e.g., a latrine may have been used once a few days ago, but flies may still be present in the latrine even though it is not frequently and consistently used by the household). Our survey was conducted during the monsoon season, and as such, data collected via latrine spot checks during this timeframe may have been somewhat influenced by seasonality. For instance, water over the latrine slab/floor may be more indicative of flooding than a lack of maintenance on the household's part. That said, a flooded latrine is perhaps less likely to be utilized. It is likely more appropriate to compare spot check indicators to the "latrine use, when" indicator, as one wouldn't expect variation in latrine users (which is what is measured with the QIS "latrine use, by whom" indicator) to vary the presence or absence of latrine spot check indicators.

4.5 Instrument-recorded use

Prior to our verification, PLUMs had never been deployed in Bangladesh, and we were uncertain of what to expect with regard to the household refusal rate. We anticipated that seeking to obtain PLUM data from slightly over one-half of QIS VWCs (14 QIS clusters, including both WASH I and WASH II QIS clusters) would result in successful collection of PLUM data from at least one-third of QIS VWCS (as requested by the RFP) after accounting for data loss (which can occur with PLUMs, particularly during monsoon season) and household level PLUM installation refusal. BRAC voiced some concern about acceptance of PLUM installation at the community level. At 6.9% (22/319), the household refusal rate amongst households in which PLUM installation was attempted was quite low. What was notable, however, were the patterns in refusal amongst wealth category. Non-poor households had the highest refusal rate, followed by poor households, and ultra-poor households. The refusal rate amongst ultra-poor households also took on a different pattern than their poor and ultra-poor counter-parts. When ultra-poor households refused PLUM installation, they often accepted installation in the primary latrine, but refused installation in the secondary latrine. These refusal patterns could perhaps have something to do with the desire for social solidarity and/or the "infectious" spread of suspicion within communities, but varying perception of the implications of refusal amongst ultra-poor households (e.g., a fear of the loss of future program benefits).

The algorithm we used to detect defecation events from verification PLUM signal data was validated in India via a comparison of PLUM-recorded signal data against structured observation of individuals utilizing their latrines (Clasen et al. 2012b). While this algorithm is validated against defecation events, it has not been validated against structured observations of child feces disposal, which may account for some level of over-reported self-reported latrine use. That said, differences between self-reported and PLUM-recorded use are similar to those seen in other contexts, with over-reporting of use observed when compared to instrument-recorded use (Clasen et al. 2012b; Thomas et al. 2013).

4.5.1 PLUM data limitations

While PLUMs are useful for capturing movement within the latrine, these remote-sensing devices do not allow for the capture of information regarding the reasons people may enter into the latrine. There may be reactivity in these data due to curiosity, children playing in the latrine, latrine superstructure materials moving/shifting during monsoon storms, etc.

4.6 Latrine utilization measurement method cost considerations

When considering the cost of the various latrine utilization measurement methodologies, it is necessary to consider the granularity and specificity of data generated by each of these measurement methodologies. While self-reported use data may be more costly than latrine spot check data, the level of detail provided in those data is considerable greater, allowing for an examination of trends in use amongst household latrine users (e.g., trends by time of day, age, and sex). While latrine spot check

indicators are cheap, they are also not very informative in quantifying use (e.g., a latrine only used by one person may not look very different from a latrine used by several people), and their level of specificity varies from indicator to indicator (e.g., a well-worn path to a latrine may not be specific to a well-used latrine, especially if the path to the latrine lies along a path to another frequently visited destination).

It is also necessary to consider the amount and types of biases accompanying each measurement method. As evidenced through our comparative analysis of self-reported and instrument-recorded use data, people tend to over-report latrine use. This over-reporting may be the result of social desirability, recall, and administration biases. Instrument-recorded data may be influenced by measurement bias, though this was minimized through the use of a validated algorithm that assesses raw PLUM signal data, and detects and quantifies signal patterns indicative of defecation events (Clasen et al. 2012). Latrine spot check data may be influenced by administration bias.

5. RECOMMENDATIONS

4. Strengthen and harmonize monitoring systems:

- **Centralize data availability:** While it is important to manage and utilize project data at field and regional office levels, it is also important to ensure data flows to, and are maintained at the central office level. We observed that high level project data (e.g., a list of all VWCs in which the project is operating, a codebook for the QIS monitoring system dataset) were not readily available at the central office. This highlighted issues with data flow and management within the project monitoring system.
 - Maintain a centralized list of all VWCs in which the WASH II project is operating. An enumerated list of all areas in which a project is operating is important not only from a program management perspective, but also a project monitoring and evaluation perspective.
 - Maintain a complete codebook of monitoring system data. The availability of a codebook will not only ensure proper interpretation of data amongst central office and M&E sub-contracting staff, but also allow future verification entities to properly analyze and interpret verification data.
- **Household identifiers:** The current monitoring system uses a different MIS and QIS household identifier for the same household. Streamlining simple aspects of the two disparate monitoring systems will serve to improve monitoring and reporting.
- **Electronic data capture:** While the MIS system electronically captures information regarding the number of latrine constructed and repaired, latrine utilization data captured on household follow-up (after receipt of hardware subsidy/loan) is not electronically entered into the database, and hardcopies of monitoring reports sent from regional offices to the central office are destroyed after three months.
- **Cross-check data prior to circulation:** One of the Excel spreadsheets containing data on intervention implementation was inaccurate – it under-reported subsidy and loan distribution because the cell formulas were not correct. In order to prevent future under- or over-reporting, databases should be cross-check for accuracy.

5. Incorporate monitoring indicators to align with the Foundation's specified outcome indicators:

- **Intervention receipt and sustained use indicators:** Consider including simple survey prompts within the QIS instrument that inquire households about when household latrines were constructed and/or last repaired (to allow program staff to determine the proportion of adults using the latrine 6 months after installation/repair), whether, when,

and by whom a subsidy or loan was received by the household for latrine construction/repair. These four prompts should add no more than 2 minutes to QIS survey administration time, but would allow program administrators to cross-check MIS data on intervention receipt, and make statements about the sustainability of latrine utilization subsequent to installation/repair.

- ***Incorporate objective indicators of latrine utilization into QIS latrine use ladder scores:*** As evidenced by the comparative analysis of self-reported and instrument-recorded data, self-reported latrine utilization data are often inflated. While it may not be practical to install sensors in latrines throughout the life of the WASH project, the inclusion of other objective indicators of latrine use may improve the veracity of self-reported latrine use.
 - Consider incorporating some of the latrine use indicators most strongly associated with use as criteria within QIS ladder score definitions.
- 6. Systematically re-assess wealth category on a routine basis:** VWC wealth category classifications are 2-7 years old, and though new households have been added to VWC registers and community maps, a systematic re-assessment of household wealth category designation has not taken place. As WASH II project intervention implementation is informed by household wealth status, it is important that the household wealth designation is accurate and up-to-date. WHO suggests that sampling frame data (the VWC register, in this circumstance) be no more than 3 years old, and advises re-enumeration of data that are greater than 3 years old.¹⁷
- 7. Encourage grantees to incorporate verifications as required activities in their proposals:**
 - ***Improve the verification process by having grantees plan for them from the outset of their projects:*** Have applicants outline how they will support verification activities, identify the line listing that the external verification entity can use as the verification sampling frame (and how often the implementer intends to update the line list), which resources they will make available to external verifying entities, and who from the implementing organization will be the main point of contact for verification inquiries.
- 8. Convene a formal stakeholder meeting (half-day workshop) prior to the initiation of future verification activities:**
 - ***Create buy-in from all stakeholders:*** Verifications are not common. The implementer may question their usefulness or may feel that they reflect a lack of confidence on the part of the donor in the implementer's own monitoring processes and reports. The implementer may not be completely informed about all donor-requested aspects of the verification exercise, and may interpret various aspects of the verification as academic exercises contrived by the verification entity rather than verification questions/objectives set forth by the donor. Before designing the verification exercise, developing a survey instrument, and collecting data, it is best for the verifying entity, donor, and implementer to participate in a half-day stakeholder workshop. It is important that the donor participate, if at all possible, in order to make clear its mandates for the purpose and scope of the verification. If it is not possible to schedule a meeting with the donor present, consider conducting the meeting in person with key staff from the verifying entity and the implementer, and invite the donor to participate via Skype or phone.
 - During the stakeholder meeting, discuss (at minimum) the following aspects of the verification exercise: 1. purpose of the verification, 2. verification scope, 3. roles of each entity and relevant points of contact, 4. list of required items, and 4. points for clarification.

¹⁷http://www.micronutrient.org/nutritiontoolkit/ModuleFolders/5.Sampling/resources/WHO_sampling_guidelines.pdf

- **Invite an array of people who represent different levels of the project:** The workshop should include participants closer to the grass roots level who are crucial to providing accurate information regarding intricate details and operational aspects of the implementer's monitoring systems and other information required to successfully conduct household visits. Future verification entities should also consider a second workshop, conducted in Bangla, with the implementer's field staff that includes role-playing techniques to help reduce fear about the verification process.

REFERENCES

Banda K, Sarkar R, Gopal S, et al. Water handling, sanitation and defecation practices in rural southern India: a knowledge, attitudes and practices study. *Trans R Soc Trop Med Hyg* 2007; **101**(11): 1124-30.

Bill & Melinda Gates Foundation. Request for Proposal: Verification Sanitation Outcomes in Bangladesh, 2013.

Clasen T, Boisson S, Routray P, et al. The effect of improved rural sanitation on diarrhoea and helminth infection: design of a cluster-randomized trial in Orissa, India. *Emerging themes in epidemiology* 2012; **9**(1): 7.

Clasen T, Fabini D, Boisson S, Song J, Aichinger E, Bui A, Dadashi S, Taneja J, Schmidt W-P, Burt Z, Nelson K (2012). Making sanitation count: developing and testing a novel device for assessing latrine use in low-income settings. *Env. Sci & Technology*. 46(6):3295-303.

Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Lippman SA, Jewell N, Bruckner T, Satariano WA. To GEE or Not to GEE. Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology*. 2010; 21:467-474

Montgomery MA, Desai MM, Elimelech M. Assessment of latrine use and quality and association with risk of trachoma in rural Tanzania. *Trans R Soc Trop Med Hyg* 2010; **104**(4): 283-9.

Sijbesma C, Bostoen K, Verhagen J. QIS monitoring guidelines for the sample study 2012. The Hague: IRC International Water and Sanitation Centre (Published November 2012).

Thomas EA, Barstow CK, Rosa G, Majorin F, Clasen T. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. *Environ Sci Technol*. 2013 Dec 3; 47(23):13602-10.