Distribution Agreement

Latrine learning: using conditional inference trees to explore how latrine conditions can predict latrine use in rural Bangladesh

By

Andrew Nute

MPH

Global Epidemiology

Thomas Clasen, JD, PhD
Committee Chair

Maryann G. Delea, MPH
Committee Member

Katie Gass, MPH, PhD
Committee Member

Latrine learning: using conditional inference trees to explore how latrine conditions can predict latrine use in rural Bangladesh

By

Andrew Nute

BA
University of Colorado
2010

Thesis Committee Chair: Thomas Clasen, JD, Ph. D.

An abstract of
a thesis submitted to the Faculty of
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2016

Abstract

Latrine learning: using conditional inference trees to explore how latrine conditions can predict latrine use in rural Bangladesh

By Andrew Nute

Global public health efforts to eliminate open defecation, specifically on the Indian subcontinent, have recently begun focusing on improving latrine use. In this study, we attempt to identify a latrine's likelihood of use based on observations of physical characteristics of the latrine and the surrounding premises (i.e., latrine spot-check indicators [SCIs]). Recursive partitioning algorithms, often called decision trees, are typically used in machine learning and data mining because they do not require the assumptions made by traditional regression models. Conditional inference trees (CIT) specifically apply unbiased statistical inference tests as a method of variable selection based on *a priori* partitioning criteria. Unlike other regression trees, the selected partitions are conditional of all other covariates in the model. In this study, we measured latrine usage in rural Bangladesh in 2014, using average daily 'likely defecation events' recorded by a motion sensing device called a passive latrine use monitor (PLUM). Using this continuous distribution, we dichotomized the measurement along its median so that we had a "most used" group (≥ median) and a "least used" group (< median). We then employed CIT to separately predict the continuous and dichotomous forms of the outcome using 15 SCIs as independent variables. After implementing a Bonferroni correction for multiple tests of significance, the CIT analysis identified a tree with three partitions using three SCIs for the dichotomous outcome. The primary partition was the presence/absence of water for the purpose of flushing or anal cleansing, with two secondary partitions being: 1) the presence/absence of flies, and 2) having a wet floor. The primary partition shows the strongest SCI but the secondary partitions show that a latrine with water for cleansing that does not attract flies and latrines that do not have water for this purpose but keep a dry floor draw the most use from their users. This interaction suggests a latrine's cleanliness and structural maintenance is an important indication of its use. The CIT for the continuous outcome could indicate some measurement error within the PLUMs.

Latrine learning: using decision tree analysis to explore how latrine conditions can predict latrine use in rural Bangladesh

By

Andrew Nute

BA
University of Colorado
2010

Thesis Committee Chair: Thomas Clasen, JD, Ph. D.

A thesis submitted to
the Faculty of Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2016

Table of Contents

BACKGROUND/LITERATURE REVIEW

Global Burden of Open Defecation

Open defecation (OD) is defined as the practice of disposing of human feces in fields, forest, bushes, open bodies of water, beaches and other open spaces (1). The Millennium Development Goals didn't directly address OD other than how it and diarrhea apply to child mortality and environmental sustainability (2). In 2015, the year by which the Millennium Development Goals (MDGs) were intended to be achieved, the MDGs missed their target for people with access to improved sanitation by 700 million people (3). In that year, 946 million people around the world practiced OD while 32% of the world's population lacked access to improved sanitation altogether. The global burden of OD is primarily shared between Africa and the Indian sub-continent.

Global Burden of Disease related to OD

Diseases related to human contact with excreta that results from OD around the world include a plethora of viruses, bacteria, protozoa and helminths (4) and very commonly lead to diarrhea, especially among children and the immunocompromised (5). Evidence suggests OD is associated with a range of public health risks, including enteric infections (6-8), malnutrition (7, 9), and other non-infectious health outcomes (10, 11). Similarly, the proportion of households found using latrines in a community is also protective against frequency of vector borne disease (12). Diarrheal diseases cause around 1.6 – 2.5 million deaths a year, a large portion of which occur among children less than 5 years of age in developing settings (13). The World Health Organization in 2014, estimated that a total of 280,000 deaths in low or middle income countries were attributable to inadequate sanitation, specifically (14).

Global Efforts to End OD

In 2015, the Sustainable Development Goals (SDGs) set by the World Health Organization and its partners, targeted OD for elimination by 2030 and set access to improved sanitation as an indicator in their monitoring plan (15). Although this is the first internationally agreed upon target set for the global elimination of OD specifically, it is a culmination of efforts to eliminate OD that preceded the creation of the SDGs. In 2000, Kamal Kar pioneered the sanitation behavior change strategy that would become known as Community-Led Total Sanitation (CLTS) while evaluating a local sanitation partner in Bangladesh (16). The approach was the first of its kind to introduce the idea of involving the community via discussion and exercise to recognize its sanitation profile and identify the resulting health impacts that it faces. This 'triggers' a sense of disgust and shame that motivates change within community members as opposed to outside influences attempting to incentivize the behavior change using a hardware subsidy to build latrines (17). Although reportedly effective, CLTS has received criticism for its use of shame and in some cases further denial of human rights related to dignity and respect (18). Regardless of this criticism, stakeholders have applied the strategy with varying levels of sustained success in many parts of the world, specifically in Africa and Asia (17, 19). Other campaigns to end OD are typically driven by governments such as India's Total Sanitation Campaign (20) where an institution that is external to the community offers the provision of individual household latrines in addition to other sanitation measures. Studies of these campaigns, however, provide evidence that improved latrine access does not always translate directly to improved latrine use (21).

OD in Bangladesh

According to *the 2015 update and MDG Assessment* published by the World Health Organization, Bangladesh reduced its rate of OD by 33% between 1990 and 2015. Although the country has made progress, there is still a considerable amount of work remaining, as over half of the population practiced OD in 2015 (3). Many different organizations have applied the CLTS strategy in different areas of Bangladesh, with a variety of modifications, since its initial design in 2000 (22). CLTS has since become a nationally implemented intervention that faces obstacles such as scalability without degradation of standards (23, 24), program variation across implementing organizations, and the associated difficulty of monitoring and evaluating these programs (25). The fundamental question to the achieving this public health goal of ending OD remains: how can implementers improve latrine use? Answering this question subsequently revolves around the more immediate task of identifying whether households are likely to use their latrine when they defecate.

Methods for Measuring Latrine Usage

Previous studies on latrine use have quantified it primarily with two different methods. The first and most common approach uses surveys and interviews to ask a respondent to accurately identify and report their use of a latrine (26). The second method is observational in nature and involves a member of the research team recording information about things he or she sees (27). Respondent-reported methods require the respondent to recall some time period and allows the researchers to inquire about any behavior that the respondent is willing to report on (i.e., defecation time, frequency and location). This method allows the researchers to sacrifice objectivity of the measure for the flexibility to obtain specific information. Studies that use direct observation are more objective in nature because the subject does not influence what information captured by the study; however, this type of study can only infer based on what can be seen which is time and resource intensive. Additionally, some studies that measure latrine use through direct observation or objective measures of use are limited in how much can be concluded because evidence of human behavior can be misinterpreted. For example, latrine spot-check indicators are characteristics of a latrine and its surrounding premises that are observed directly by a researcher, but are not directly indicative of the latrine's use.

In 2012, Clasen et al. (28) introduced a device called a passive latrine use monitor or PLUM device. The device uses a passive infrared motion detector similar to those used in industrial complexes for detecting human activity. The device records binary signal data as it detects a specified infrared wavelength consistent with that of human skin. These signal data are then interpreted by collaborators within the Sustainable Water, Energy and Environmental Technologies Laboratory (SWEETlab) at Portland State University. The SWEETlab researchers apply a validated algorithm that scans the data and detects patterns that are indicative of 'likely defecation events.'

Biases Affecting Each Measuring Method

Objectively measuring a subject's behavior without affecting that behavior is a challenge to researchers, especially when it is a private act such as latrine use. Both survey and observational methods are impacted by bias in the form of social desirability or courtesy bias. This results from a subject wanting to report information that he or she thinks is more acceptable to others, specifically the people he or she is reporting to. Market research has demonstrated this bias (29), as have behavioral studies, including latrine use (30) in India. This bias was also shown to exist in studies using direct observation to measure latrine use, such as Clasen et al. in 2012 (28), during which the authors introduced the PLUM device.

Although this device is highly capable of detecting human motion and continued presence, recordings and associated data are nonetheless commonly affected by changes in subjects' behavior due to their knowledge of the recording device's presence. This was termed reactivity and evidence from cook stoves and water filter use monitors (31) that also operate with automated monitoring devices suggest it does, indeed, exist. Other studies commonly term this biasing effect the Hawthorne effect and have quantified it with the use of devices that improve hand hygiene (32).

Application of Recursive Partitioning Methods

Recursive partitioning is the act of splitting a predictor variable at every possible point in its distribution to see which "partition" provides the most desired grouping in a response variable according to standards determined by the user and the algorithm used (33). The resulting graph after multiple variables are partitioned upon is called a decision tree which starts with a root node and ends in leaf nodes also referred to as terminal nodes (34). Each node that is split is commonly termed a parent of the two resulting nodes which are conversely termed child nodes of that parent node (35).

Through the employment of a recursive partitioning algorithm called Conditional Inference Trees (CIT), this paper attempts to identify latrine spot-check indicators that are most associated with latrine use. Decision trees are often used in data mining and machine learning techniques (36, 37) to find patterns in data (38). Decision trees allow users to determine measures of association between the predictor and response variables if the user is able to assume the direction of the association (39) and apply epidemiologic concepts for determining risk or odds, as necessitated by the study design. The decision tree, however, does not give a parameter for this measurement like a regression coefficient in logistic or linear modeling. Some decision trees built using certain recursive partitioning algorithms like CIT can, however, show statistical significance of partitions included in the decision tree (40). More details on specifically how CIT determines partitions and differs from other decision tree algorithms will be discussed in the methods below.

Researchers use decision trees, specifically Classification and Regression Trees (CART), to segment populations into subgroups with differing likelihoods of engaging in behaviors that are determinants of certain health outcomes (41). A major benefit of using a decision tree for classification purposes is that an implementer can determine the likelihood of modeled categorical outcomes without the need for calculations or complex measurement indices (42). This paper attempts to use classification decision trees to enable someone to determine the likelihood that a latrine falls above or below the median use per user based on latrine spot-check indicators.

We are also interested to see which indicators would provide the most statistically significant partitions of the continuous outcomes from which the ordinal variables were created. CIT uses a form of the T test to compare the means of the two distributions of a continuous response variable within each child node of a partition (43). The shape of the underlying distributions being tested does not hinder the T test's ability to differentiate them with a test of significance, in the case of this paper, both respondent-reported and PLUM-recorded latrine use have negative binomial distributions. This is evidenced by the central limit theorem which proves that the mean of a distribution will be normally distributed regardless of the distribution of the variable from which it is derived (44). Although less useful in the field, using the CIT algorithm with a continuous outcome does not force the user to transform a continuous response variable like latrine use into categories. The subsequent fidelity to the response variable's true distribution allows the CIT algorithm to apply a significance test that is more appropriate for the measurement of interest. The resulting model gives a better indication of which variables provide the partition with the strongest difference in the child distributions.

Although classification and regression trees (CART), as developed by Breiman et al. (33), are among the most commonly used of these partitioning methods, CART lacks the ability to account for distributive properties among variables which, in turn, causes the method to favor variables with high degrees of missingness (45). This would present a problem in this data only if we wanted to include our last indicator (availability of a cleansing agent for hand washing within one minute) as a possible partition because it has fewer responses than the other indicators. CART would favor this indicator because of its smaller response rate. If we excluded this indicator, CART would remain unbiased because it is the only indicator with fewer observations than the others.

Unlike CIT, CART builds its model by starting with a tree that is over-fit to the data and requires the investigator to “prune” the tree based on a complexity parameter which is determined by the researcher who considers an acceptable misclassification cost within separate terminal nodes of the model (33). This process often leads to models that lack parsimony and arguably subjectively fit the data to meet the author’s needs and not necessarily the needs of future users of the model. Hothorn et al. offered CIT (43) as an alternative recursive partitioning algorithm that uses significance tests to compare variables and rank them within the tree rather than allowing for subjective and parameterized selection processes. We decided to use CIT because it will consider all variables in a non-parameteric way and will objectively decide which spot-check indicator provides the most statistically significant partitions, and should consequently be given priority. This process also prevents issues arising from multicollinearity in the data, which is a common analytical issue when traditional regression models are used to analyze correlated independent variables (46). A step by step explanation for the variable selection process within CIT in this analysis is included in the methods section found below. As a newer method of recursive partitioning, researchers have used CIT in a number of situations both related and unrelated to health. More recently, researchers have suggested its use for exploring various WASH topics (47) and topics of behavioral outcomes such as intimate partner violence (48).

METHODS

Objectives

The primary objective of this paper is to provide persons in the field with a model which can be used to predict ordinal levels of latrine usage (e.g., above or below the median) in rural Bangladesh based on observed latrine spot-check indicators. Latrine use in this paper is measured in two ways: 1) recall of the general number of times per day users of the household latrine used the facility (i.e., respondent-reported latrine use) 2) passive latrine use monitor (PLUM)-recorded events determined to be 'likely defecation events' (i.e., PLUM-recorded use). The secondary objective of this paper is to determine which, if any, of the indicators can provide the strongest partition when considering these two outcome measures on a continuous scale. If any indicators are found significant after Bonferroni correction, we will investigate the differences in the subsequent distributions for their functionality in the field.

Research Question 1: Using a recursive partitioning method known as Conditional Inference Trees (CIT), can latrine spot-check data be used to predict categorized respondent-reported or PLUM-recorded latrine use?

$H0_1$: None of the 15 independent latrine spot-check variables will provide a significant partition at alpha: 0.05 after using Bonferroni correction for multiple tests of significance.

$H1_1$: CIT will identify at least one out of the 15 independent latrine spot-check variables as significant (P-value < 0.05) after using Bonferroni correction in predicting ordinal categories of latrine usage in Bangladesh as determined by above or below the median for either respondent-reported or PLUM-recorded 'likely defecation events'.

<u>Research Question 2:</u> Using Conditional Inference Trees, can latrine spot-check data be used to determine statistically significant differences in continuous distributions of respondent-reported or PLUM-recorded 'likely defecation events'?

$H0_2$: None of the 15 independent latrine spot-check variables will provide a significant partition at alpha: 0.05 after using Bonferroni correction for multiple tests of significance.

$H1_2$: CIT will identify at least one out of the 15 independent latrine spot-check variables as significant (P-value < 0.05) after using Bonferroni correction in differentiating continuous distributions of respondent-reported or PLUM-recorded 'likely defecation events'.

Data Source

In 2014, The Bill and Melinda Gates Foundation (The Foundation) commissioned Emory University to execute a sanitation outcome verification study. Emory University subsequently partnered with the International Centre for Diarrhoeal Disease Research, Bangladesh and Portland State University to design and coordinate the verification. The study was intended to verify programmatic results reported by a non-governmental organization (NGO) that received a grant from The Foundation to implement large-scale Water, Sanitation and Hygiene (WASH) projects in Bangladesh (49). One objective of the study was to compare various latrine use measurement methods: respondent-reported, PLUM-recorded, and latrine spot check indicators of use.

After employing a Monte Carlo simulation to conduct a sample size determination in SAS software version 9.4, participating households were selected into the study using a multi-stage sampling strategy. This strategy selected village-WASH committee (VWC) clusters in which the NGO was implementing its Foundation-funded WASH projects. The intention of the sampling process was to test two-sided verification hypotheses to determine whether the sanitation outcomes reported by the NGO were reliable. In the initial sampling stages, study staff randomly selected, with probability proportionate to size, 26 sampling units from each of two WASH intervention groups, for a total of 52 sampling units. Subsequently, study staff randomly selected one VWC cluster from each sampling unit for inclusion into the study using simple random sampling. In the final round of sampling, the field team obtained the register of households in each selected VWC, and stratified the households by wealth category (as defined by the VWC and the NGO). Study staff employed systematic random sampling to selected eight households from each of the three wealth categories (ultra-poor, poor and non-poor). The study design and sampling methods are described in greater detail in the verification report provided to The Foundation after the study was completed (49).

For the purposes of the larger verification study, a household was 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 the head of household.” (50) Households were excluded after selection if they refused participation and/or consent, were absent all three times the survey team visited the household, or did not have an adult (aged 18+) at home to serve as the survey respondent.

In order to obtain household latrine use measures, the study team performed household surveys, which consisted of the administration of a questionnaire and spot-checks of household latrines and the surrounding premises (performed on all household latrines) amongst all households randomly selected for inclusion in the study during June-August (i.e., monsoon season) 2014. In addition to these measures, the study team randomly selected a sub-set of study clusters in which PLUM devices were installed in all household latrines.

Study Population

The resulting verification study population consisted of 1207 households from rural Bangladesh, of which 213 households contributed PLUM data for the final analytical sample. Figure 1 (51) shows the flow of PLUM data capture and gives a rationale for why households selected for inclusion in the PLUM sub-sample were excluded from our final analytical sample.

During the course of analysis, it was determined that households with more than one household latrine should be dropped from the analysis for this study because the latrine spot-check indicators could not be directly related to the latrine's use by the household. Figure 2 shows the flow of households that were excluded in this study from both the PLUM and overall study population. Of the 1207 households, 1191 indicated that their most commonly used latrine was functional. Of these 1191 households, 102 indicated that there is a second functional latrine on the household compound and 13 indicated they had a second latrine but were missing information about its functionality. This brought the total analytical sample for this study down to 1076 households. A total of 197 of these households contributed PLUM data and collectively make up the analytical sample for the PLUM outcome models. Table 1 shows descriptive statistics for the 1076 households included in models predicting respondent-reported daily counts of latrine use and the 197 households included in models predicting PLUM-recorded use per day. Table 2 displays the spot-check indicators within the two groups.

Bias analysis for PLUM Devices vs Respondent-Reported Events:

As previously mentioned, there was an expected amount of reactivity in latrine usage among households to the installation of the PLUMs. This reactivity was expected to be much stronger during the beginning of the seven day installation period which was planned for each household. As a result, it was decided before installation that the first two days of data recorded by each PLUM would be excluded from the analysis as the study team presumed data from those days were likely the most affected. Although some of the PLUM devices recorded more days than expected due to logistical complications during the study (e.g., breaks for Eid celebration), only days three through six of the PLUM-installation period will be included in the analysis for consistency across PLUM devices and to keep the recorded times as close as possible to the time of the spot-check data but after the proposed reactivity period. An analysis of respondent reported and the PLUM recorded use measures was performed on these data (50) to provide insight on the potential for bias between measurement methods.

Dichotomizing the Outcomes

After a thorough search of available literature, we found no parameter that could be applied to this study population that reflected the average number of defecation events a person has per day. Studies have identified the average number of daily defecation events per person in other populations in western settings (52) but those studies could lack the external validity needed to be applicable to this study's comparatively less wealthy population which also has a much different diet. In an attempt to create the most useful models for an implementer in the field based on the two outcomes of interest, a cut point would be needed to create classifications of use for each latrine. Taking into consideration the negative binomial distribution of both of the outcomes of interest, the median of each outcome was used to dichotomize the sample of latrines into two groups, "most used" being those that were equal to or above the median and "least used" being those that were below the median. The median for respondent-reported use counts (per latrine user per day) among the 1076 latrines is 1.25. The median number of the PLUM-recorded events among the 197 households in this study was 0.81.

Statistical Methods:

All CITs were built within R Studio (53) using the party package (43, 54) within the R programming language (55).

As a result of having a large number of significance tests for multiple independent variables, a Bonferroni correction was implemented to reduce the subsequent increased possibility of type I error (56, 57). After this correction was applied to significance tests, a significance level (α) of 0.05 was used to determine whether a partition was significant. This was done because all the independent variables could intuitively be an indication of latrine use but none of them was hypothesized to be the most relevant prior to building the CIT. In addition, an analysis of correlations between spot-check indicators was performed to give a sense of which variables might tell us the same information. It should be noted that this is only intended to frame the indicators for anyone using the resulting model. As mentioned before, the CIT algorithm is not affected by multicollinearity among the independent variables (46). Table 3 shows the correlations between these indicators.

We also grew decision trees using CIT with univariate significance testing for comparison to their Bonferroni-corrected counterparts. This was done for the purpose of observing which partitions could be unstable and are therefore eliminated. A number of reasons can lead to a variable being unstable such as a very small effect size of a particular partition or a small number of observations that fit into the terminal nodes after the stratification that results from each partitioning level. These trees are not reliable and are only intended to frame their Bonferroni-corrected counterparts. These trees were also grown because they can provide evidence that could warrant further data collection.

Acceptable Partition Criteria

Three criteria had to be met for any partition to be allowed into the models in Figures 3 - 7. First, we used a minimum split criterion of twenty observations which means that any node could not be further partitioned if there were fewer than twenty observations in that node. Second, we used a minimum terminal node criterion of five observations meaning that a partition was prevented if one of the resulting child nodes did not consist of at least five observations. Finally, if a partition's test statistic was not significant at the significance level (α) stated above after Bonferroni correction, then that partition was prevented.

Determining Significance of Partitions

It is important to realize that all of the independent variables used to predict the outcomes in this paper are binary variables. This means that there was only one possible partition (Yes vs No) that could be fit into each variable simplifying the recursive partitioning process because the CIT algorithm does not continuously search for the best split in each predictor. If we had a continuous predictor variable being used, the CIT algorithm would split that variable at every possible point and find the most significant partition provided in the response variable by that predictor variable. This does not happen in this analysis, however.

In Figures 3, 4 and 6, the model considers the outcome as an ordinal variable and compares a matrix of the binary dependent variable and independent variable. The resulting test of significance is a χ^2 test of independence and corresponds to a χ^2 test statistic (43).

In Figures 5 and 7, the CIT algorithm treated the outcome as a continuous variable, and the statistic being tested is close to but not exactly a true two sample T test. The applied T statistic is primarily different in how it adjusts for differences in sample size and misclassification error. For specifics on how this is derived, please see Hothorn et al. (43)

Variable Selection in CIT

In all of these models, at the start of the process, the algorithm performs a significance test for each predictor's influence on the outcome being modeled. At this point, if the most significant partition is still significant at the pre-determined level for α after Bonferroni correction, it is applied as the primary partition for the model. This process is then repeated on both child nodes with the remaining predictor variables until all partitioning is stopped by one of the three partitioning criteria mentioned above.

Ethical Considerations and Approval

Prior to conducting the primary sanitation verification study, Emory's Institutional Review Board granted this study approval (IRB00073752) as did the Ethical Review committee of the International Centre for Diarrhoeal Disease Research, Bangladesh. All ethical considerations for this study should be referred to within the larger validation study, as this paper represents a secondary analysis performed on de-identified data.

RESULTS

Correlation of Spot-check Indicators

The presence of fecal or urine odor was found to be correlated with traces of feces on the latrine pan or slab. Additionally, the availability of water for handwashing was found to be correlated with the availability of water for flushing or anal-cleansing.

PLUM-recorded 'likely defecation events' Averaged across Latrine Users

After applying the cut point of 0.81 derived from the median PLUM recorded likely defecation events and implementing Bonferroni correction, the decision tree in Figure 3 was grown using CIT. This shows the primary variable of importance to be the availability of water for the purpose of flushing or anal-cleansing (p-value: 0.012) with two secondary partitions splitting based on the presence of flies (p-value: 0.005) and a wet floor (p-value: 0.037). The four terminal nodes show two pairs of terminal nodes showing nearly a direct contrast between least used and most used groups. Among latrines where there is water present, the presence of flies indicates a little more than 80% probability of the latrine falling into the least used group while the absence of flies indicates almost an 80% probability that a latrine falls into the most used group of latrines. Among households that do not have water available for flushing or anal-cleansing, the presence of a wet floor indicates close to a 65% probability that the latrine belongs to the least used group while a dry floor shows almost the exact same likelihood of being in the opposite, most used group.

The CIT plot for this outcome that did not correct for multiple tests of significance shown in Figure 4, identified the primary partition of importance to be the availability of water for flushing of anal-cleansing along with the secondary partitions: 1) presence/absence of an odor emanating from feces or urine, and 2) a wet floor.

Respondent-reported Daily Events Averaged across Latrine Users

After applying the previously determined median cut point of 1.25 daily likely defecation events per latrine user and implementing Bonferroni correction, the CIT algorithm did not identify any significant partitions. Figure 6 shows the tree that resulted without correcting for multiple tests of significance. Although a number of possible splits are identified, they could have been a chance result and would need further investigation using new data to be considered actually predictive of the ordinal outcome of least used or most used groups.

CITs for Continuous Outcomes Showing Strongest Partitions

Figure 5 shows the CIT that selects the strongest partition while predicting the PLUM-recorded outcome on a continuous scale. The strongest partition is provided by the indicator that asks about the presence/absence of water for the purpose of flushing or anal-cleansing. Figure 5a shows the kernel density estimations (aka a non-parametric estimate of the density distribution) for each of the two distributions in the terminal nodes of Figure 5. The two estimations are overlaid to display their differences that led to their inclusion as the primary partition in the CIT in Figure 5. Figure 5b shows the kernel density estimations from the two distributions that would have resulted from substituting the indicator chosen by CIT with a second indicator, presence/absence of water for handwashing which was highly correlated with the indicator chosen in Figure 5.

Figure 7 shows the CIT that selects the strongest partition while predicting the respondent-reported outcome on a continuous scale. The strongest partition is provided by the indicator that asks about the presence/absence of water for the purpose of handwashing. Figure 7a shows the kernel density estimations for each of the two distributions in the terminal nodes of Figure 7. Similar to Figure 5a, the two estimations are overlaid to display their differences that led to their inclusion in the CIT shown in Figure 7. Figure 7b shows the kernel density estimations from the two distributions that would have resulted from substituting the indicator chosen by CIT with a second indicator, presence/absence of water for flushing or anal-cleansing which was highly correlated with the indicator chosen in Figure 7.

DISCUSSION

I. Purpose

The purpose of this analysis was to: 1) see if decision trees created with the CIT algorithm could build a model for implementers to use in the field, and 2) identify which of the latrine spot-check indicators provided the most significant test statistic while partitioning the study population. The difficulty associated with determining latrine usage presents major challenges to those organizations that would attempt to improve it. Using de-identified data with two different measures of latrine use, we built the models by employing a data mining technique to search for patterns in the data and test their significance.

The application of decision trees to the field of health (41), human behavior (48) and WASH (47) are not new concepts. This paper is novel in that it specifically looks at using these techniques to explain a human behavior relating to WASH that could impact human health. Machine learning and data mining techniques like this forego reasoning beyond algorithmic calculations. By doing this, the global health community could start to see patterns through data that are not apparent otherwise. In this circumstance, using these techniques could lead to understanding driving factors for choosing to use a latrine as opposed to practicing OD.

II. Findings

Interpreting Correlated Latrine Spot-check Indicators

The correlation between the presence of fecal odor and traces of feces is understandable considering the odor could emanate from the feces. Similarly, water for either the purpose of washing one's hands or for the purpose of flushing the toilet are likely to come from the same source and so the correlation between the two could also be expected in a setting like rural Bangladesh where access to water is not universal. Although these correlations will not have an effect on the predictive models, they could be considered if someone attempted to use these models in the field and one of the correlated indicators was missing.

Categorical Response Variable

When one parent partition leads to a child partition, it signifies an interaction of the two that allows us to infer things about the resulting sub-populations. For example, the results in Figure 3 show an interaction between the primary and secondary nodes that, together, could be an indication that a latrine with a high degree of hygienic cleanliness and structural functionality is one that is likely to be used. In latrines that have water for cleansing/flushing, flies being absent could be a sign that people are using the latrine and using the water to remove the feces that would attract the flies. Latrines that do not have water for flushing/anal-cleansing and yet have a wet floor could indicate a structure that is unable to keep water out of the latrine that was not intended to be there (e.g., rain or flooding). A closer look at roof types in this sub-population shows that of the 96 latrines in Node 7 of Figure 3, 6% have an improved roof (i.e., tiles or concrete) compared to the 13% in the 197 households included in the total population for this model.

Comparing Bonferroni corrected trees to trees that are uncorrected in this manner offers insight into which partitions might be unstable. For instance, Figures 3 and 4 are nearly identical with the exception of the 2nd, 3rd and 4th nodes, which together form the parent and child nodes involved in the partitioning based on the presence/absence of an odor emanating from the latrine in Figure 4 instead of the presence of flies shown in Figure 3. This is likely a result of having a sample size in the terminal Nodes 3 and 4 of Figure 4 (n = 40, n = 20) that was considerably more balanced than those provided by the presence of flies in Figure 3 (n = 55, n = 8). The presence of an odor, however, provided a much weaker predictive split in Node 4 of Figure 4 compared to Node 4 of Figure 3, which makes it much less stable despite the larger sample size in the terminal node. The sparse data in Node 4 of Figure 3 is also cause for concern in terms of that spot-check indicator's predictive value within a larger sample of latrines meeting this description. If there were larger amounts of latrines in this node, the presence of flies could turn out to be less effective of a predictor and could even be eliminated by the higher significance standards that eliminated the presence of odor as a useful partition.

Continuous Response Variable and Kernel Density Estimations:

The single partition shown in Figure 5 uses the same latrine spot-check indicator as the primary partition shown in both Figures 3 and 4. Similarly, the single partition in Figure 7 uses the same latrine spot-check indicator as the primary partition in Figure 6. These two facts help to validate the categorical decision trees as providing partitions as a result of the indicators' significance of association with the outcome and not providing the partitions as a result of the transformation of the outcome from continuous to categorical. Figures 7a and 7b show these two latrine spot check indicators, the presence/absence of water for flushing or anal-cleansing and the presence/absence of water for handwashing, to have nearly identical kernel density estimations for their respectively dichotomized responses. This means the two indicators offer roughly the same predictive value for respondent-reported latrine use that would be portrayed in a partition of a decision tree. In Figures 5a and 5b, however, we do not see this similarity between the two correlated latrine spot-check indicators. Figure 5a shows a much more drastic difference between the two kernel density estimations provided by the presence/absence of water for flushing or anal-cleansing than the difference in the two kernel density estimations provided in Figure 5b. It is worth noting that the act of flushing or anal-cleansing would involve a motion that the PLUM device would be likely to detect. Furthermore, the PLUM device would detect this activity more than the act of handwashing which could even take place outside of the latrine. This raises the question, could the PLUM device be more sensitive to this variable because the PLUM device is programmed to detect human motion. In other words, could there be some measurement error in the form of a lack of perfect specificity in the PLUM-recorded data?

III. Implications

The findings of this analysis indicate that reasons exist for people to practice open defecation other than poor access to a latrine because not all people that have access to a latrine use that latrine. This adds more evidence to the hypothesis that latrine coverage or a lack thereof, is not the primary influence for a community's rate of open defecation. We attempted to see if latrine spot-check indicators can predict latrine use but found that, how latrine use is measured is important in determining if that is possible. Stakeholders and governments that look to employ these decision trees in the field should be aware of their strengths and limitations. Future researchers looking to use PLUM devices as a method of measuring latrine use should be aware that as motion sensors, their specificity for identifying 'likely defecation events' is potentially less than perfect. As a result, they could register false positives for 'likely defecation events'.

STRENGTHS AND WEAKNESSES

Strengths

The population size of the primary study provided a large enough sample size for the CIT algorithm to find significant partitions in the sub-population of households with latrines with PLUM devices installed. We were also able to use Bonferroni-correction in the CIT algorithm which strengthens the evidence that the results we found were not a result of chance. Furthermore, the use of CIT as our partitioning algorithm provided an objective method to grow the decision trees. As a novice decision tree grower, we recognized this benefit that was not shared by the more established recursive partitioning method CART.

Weaknesses

The primary limitation to this study is that we cannot properly label a latrine as an always, sometimes or never used latrine. Doing this would provide a much more relevant conclusion to policy makers as households could then be identified as target beneficiaries of a behavior change communication strategy that was more focused than the community level. It is possible that a latrine in this study could be used more because of other factors unrelated to the latrine such as the demographics of its users. Although we have this demographic information, we need to find and apply external parameters in order to interpret how the demographic information will affect our expectations of how much each latrine should be used.

As part of the study, data exclusion and manipulation was a necessary step to reduce bias and improve comparability between our two measures. Although the PLUM devices were installed for more than a four day period, we decided to exclude the first two days because of the reactivity to the devices that we expected to see among latrine users. It is possible, however, that there was no reactivity and as a result, we excluded perfectly fine data that would have allowed for more accurate estimations of PLUM-recorded latrine use. We also averaged the PLUM-recorded 'likely defecation events' to get an estimate on a "per user, per day" scale that would be comparable between the two measurement methods.

Finally, as a predictive model, another major weakness of this analysis is that when CIT does find significant variables, it still cannot determine directionality of the associations between the predictor variables and the response variable. The users have to interpret this directionality externally to the created decision tree. For some of the latrine spot-check indicators, we cannot assume for certain that the predictor is causing the response variable rather than the response variable being a cause for the predictor. It is difficult, for example, to say whether the absence of water for flushing the toilet or anal-cleansing is a reason that a person decides not to use the latrine or if there is no water for this purpose because that latrine is so rarely used and the water would serve no function.

FUTURE RESEARCH AND DIRECTIONS

Measurement of PLUM-recorded Latrine Use

More research is needed to understand if the PLUM devices have a propensity to detect more 'likely defecation events' in latrines where people are likely to cleanse themselves with available water. This seems possible according to the results of using CIT to predict PLUM-recorded events on a continuous scale. The PLUM device's objectivity makes it a great tool for researchers of latrine use around the world. Despite this fact, they need continued improvement to meet the needs for measurement precision that would make them the gold-standard for measuring latrine use.

Measuring Latrine Use

Despite their potential for a lack of perfect specificity, we suggest using the PLUM-recorded measure of latrine use be used as opposed to respondent-reported counts. This is because of the objectivity of the measure compared to the respondent-reported measure. Although there is need for improvement, at least there is the ability to improve this measure and adjust for biases which is a much more difficult task for the respondent-reported measure after it is collected.

Future for Policy

The search for estimations of defecation frequency among this population needs to continue. If parameters can be applied to the latrine users for each latrine, we could then estimate each latrine's expected use. This would be relevant for policy because each latrine could then be considered "Always" "Sometimes" or "Never" used. This would help implementing organizations to identify people that practice OD allowing intervention strategies to be more targeted to these individuals.

Other Decision Tree Algorithms

Recursive partitioning and their resulting decision tree models are used by a number of algorithms. Although first used by Breiman et al. (33) in CART, other algorithms have become popular in the scientific community. We suggest that future papers using this data employ other methods such as CART or Random Forests (58) as the results may be different.

TABLES AND FIGURES

Figure 1. Flow of PLUM data capture from Latrine Manuscript Outline (51)

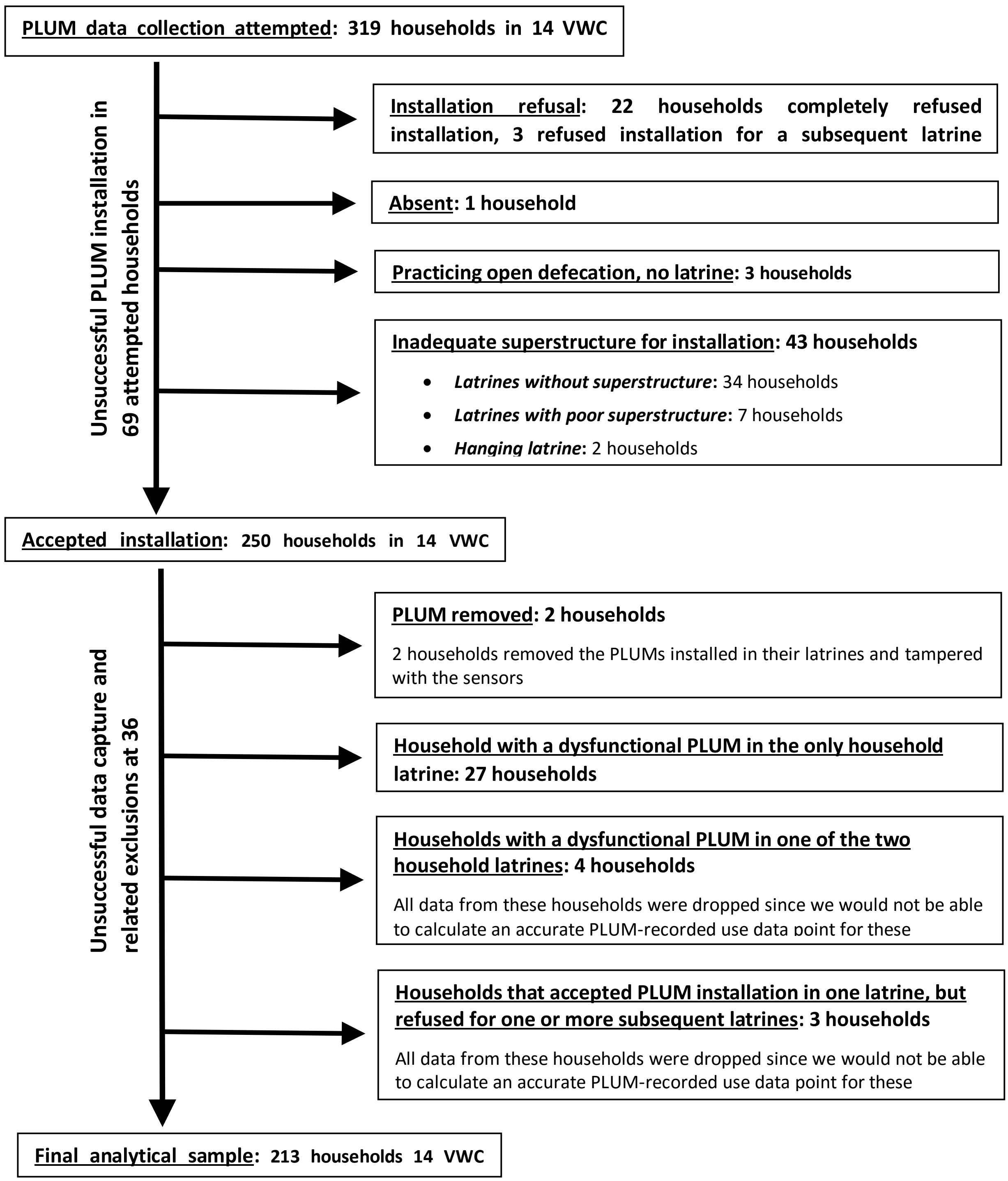


Figure 2. Flow of Household Exclusion from CIT models

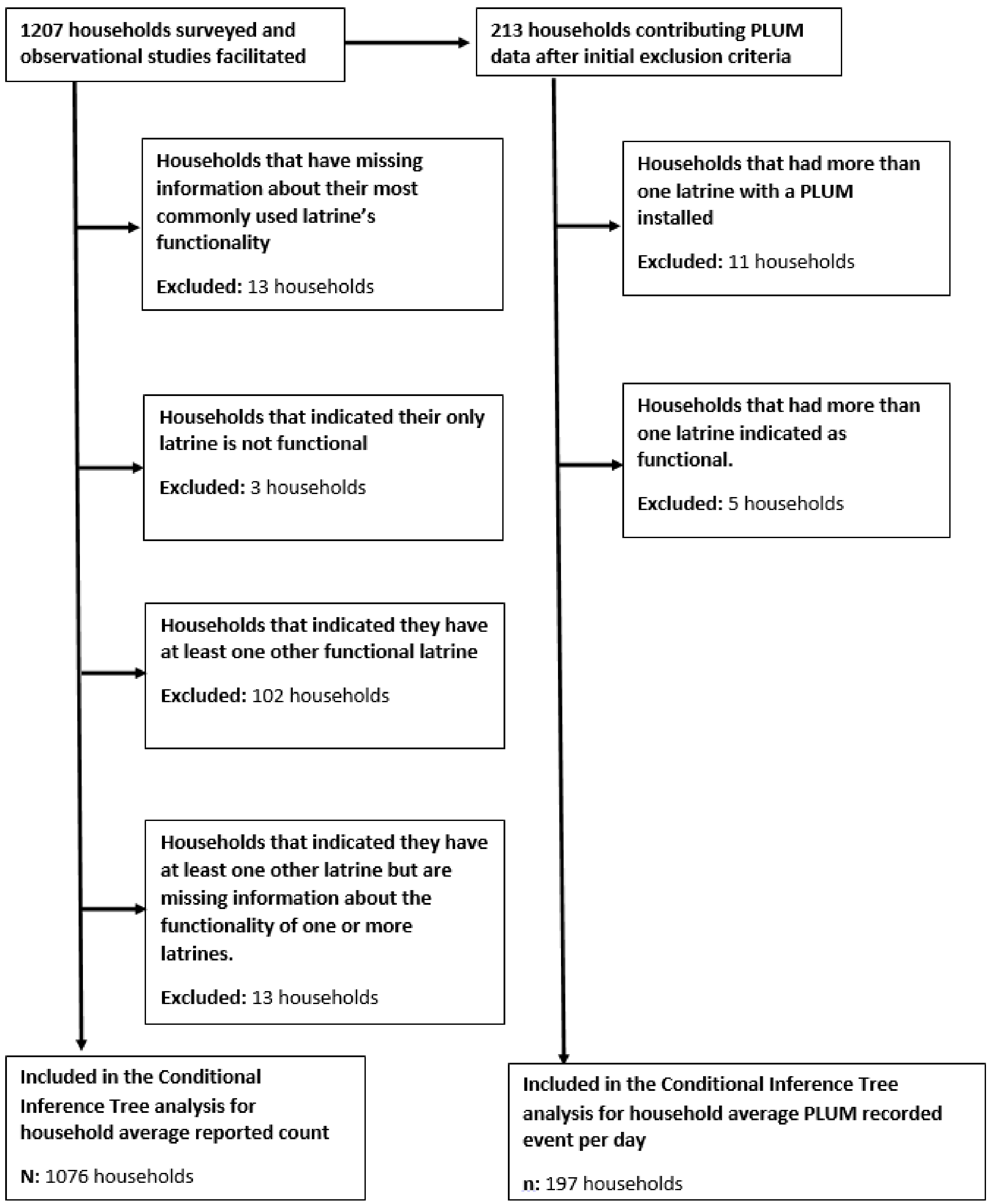


Table 1. Sample Characteristics

	All HH in the study population (N = 1076)		Study HH with a funtional PLUM device (n = 197)	
Household characteristics	**Median**	**IQR**	**Median**	**IQR**
Number of household members (>3 years)	5	(4, 7)	5	(4, 7)
Number of latrine users in each household	5	(4, 7)	5	(4, 7)
Age of respondent	35	(26, 45)	35	7, 45)
Respondent's years of formal education	4	(0, 5)	4	(0, 7)
	N	**(%)**	**n**	**(%)**
Household Wealth Category				
Non-poor	354	(33%)	70	(36%)
Poor	346	(32%)	58	(29%)
Ultra Poor	376	(35%)	69	(35%)
Sex of the respondent				
Male	112	(10%)	31	(16%)
Female	964	(90%)	166	(84%)
Latrine characteristics	**N**	**(%)**	**n**	**(%)**
Types of latrines in the sample				
Flush to pit	328	(30%)	51	(26%)
Flush to Septic tank	47	(4%)	14	(7%)
Piped sewer system/ Flush to elsewhere	5	(< 1%)	3	(2%)
Pit latrine with slab & water seal	306	(28%)	69	(35%)
Ventilated Improved Pit latine (VIP)	5	(< 1%)	1	(1%)
Open pit latrine	381	(35%)	58	(29%)
Hanging toilet/latrine	4	(< 1%)	1	(1%)
Materials for latrine enclosure*				
Plastered bricks with tiles	6	(< 1%)	0	
Plastered bricks	229	(21%)	34	(17%)
Un-plastered bricks	39	(4%)	10	(5%)
Tin/metal	398	(37%)	95	(48%)
Bamboo/wood	135	(13%)	18	(9%)
Cloth/plastic/sack	186	(17%)	23	(12%)
Other	74	(7%)	16	(8%)
Materials used for door**				
Tin/metal	496	(46%)	109	(55%)
Plastic sheet	141	(13%)	19	(10%)
Cloth/curtain	296	(28%)	46	(23%)
Bamboo/wood	66	(6%)	12	(6%)
No door	34	(3%)	3	(2%)
Other	18	(2%)	4	(2%)
Materials used for roof				
Concrete	162	(15%)	20	(10%)
Tiles/tally	14	(1%)	5	(3%)
Corrugated tin	490	(46%)	119	(60%)
Thatch/grass/plastic	96	(9%)	20	(10%)
No roof	310	(29%)	33	(17%)
Other	4	(< 1%)	0	

*9/1076 and 1/197 missing information about enclosure material

**25/1076 and 4/197 respectively missing information about door material

Table 2. Distribution of Spot-check Variables

Spot check indicators	Total households included in the study (N = 1076)				Housholds with a PLUM and only 1 latrine (n = 197)			
	No	(%)	Yes	(%)	No	(%)	Yes	(%)
1. Latrine being used for storage?	1075	(100%)	1	(<1%)	197	(100%)	0	(0%)
2. Stagnant water on floor?*	959	(89%)	113	(11%)	180	(92%)	16	(8%)
3. Traces of feces on latrine pan/slab?*	487	(45%)	586	(55%)	104	(53%)	92	(47%)
4. Discoloration of pan/slab?*	446	(42%)	625	(58%)	89	(45%)	107	(55%)
5. Presence of fecal/urine odor?	407	(38%)	669	(62%)	82	(42%)	115	(58%)
6. Presence of spider webs, leaves or other debris indicating lack of use?	789	(73%)	287	(27%)	138	(70%)	59	(30%)
7. Presence of flies inside?	684	(64%)	392	(36%)	137	(70%)	60	(30%)
8. Availability of cleaning agents for washing the latrine?	922	(86%)	154	(14%)	176	(89%)	21	(11%)
9. Presence of well-worn path?	50	(5%)	1026	(95%)	10	(5%)	187	(95%)
10. Is the floor wet?*	291	(27%)	783	(73%)	57	(29%)	140	(71%)
11. Presence of slippers outside?	879	(82%)	197	(18%)	164	(83%)	33	(17%)
12. Availability of water for flushing/self cleansing?	759	(71%)	317	(29%)	134	(68%)	63	(32%)
13. Availability of water for hand washing?	701	(65%)	375	(35%)	116	(59%)	81	(41%)
14. Availability of cleansing agent near or inside?	743	(69%)	333	(31%)	142	(72%)	55	(28%)
15. Ability to obtain cleansing agent within 1 minute?**	166	(22%)	577	(78%)	24	(17%)	118	(83%)

* Variable is missing in ≤ 5 latrines for one or both latrine use measures

** Variable is missing in 333/1076 and 55/197 latrines

Table 3. Pearson's Correlation between all 15 spot-check variables with p-values excluded. Only values >0.6 were typed in bold.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Latrine being used for storage?	1														
2. Stagnant water on floor?	0.081	1													
3. Traces of feces on latrine pan/slab?	0.027	0.062	1												
4. Discoloration of pan/slab?	-0.033	0.053	0.437	1											
5. Presence of fecal/urine odor?	0.024	-0.008	**0.653**	0.482	1										
6. Presence of spider webs, leaves or other debris indicating lack of use?	-0.017	0.045	0.279	0.219	0.274	1									
7. Presence of flies inside?	-0.022	-0.125	0.437	0.336	0.556	0.144	1								
8. Availability of cleaning agents for washing the latrine?	-0.013	0.044	-0.276	-0.270	-0.329	-0.130	-0.247	1							
9. Presence of well-worn path?	0.006	0.027	0.039	-0.014	0.029	0.027	-0.015	0.052	1						
10. Is the floor wet?	0.018	0.180	0.107	0.101	0.023	0.045	-0.022	-0.026	0.032	1					
11. Presence of slippers outside?	-0.014	-0.003	-0.238	-0.272	-0.326	-0.132	-0.218	0.301	0.098	0.014	1				
12. Availability of water for flushing/self cleansing?	0.043	0.064	-0.250	-0.180	-0.343	-0.197	-0.306	0.381	0.042	0.022	0.323	1			
13. Availability of water for hand washing?	0.038	0.051	-0.262	-0.134	-0.317	-0.201	-0.255	0.343	0.039	-0.039	0.279	**0.826**	1		
14. Availability of cleansing agent near or inside?	-0.020	0.014	-0.287	-0.278	-0.405	-0.219	-0.319	0.399	0.039	-0.006	0.464	0.522	0.495	1	
15. Ability to obtain cleansing agent within 1	0.018	-0.111	-0.148	-0.150	-0.169	-0.043	-0.159	0.106	-0.051	0.088	0.058	0.109	0.060	.	1

Figure 3. CIT Plot for PLUM recorded use per latrine user categorized using the median cut point (0.81) with Bonferroni correction.

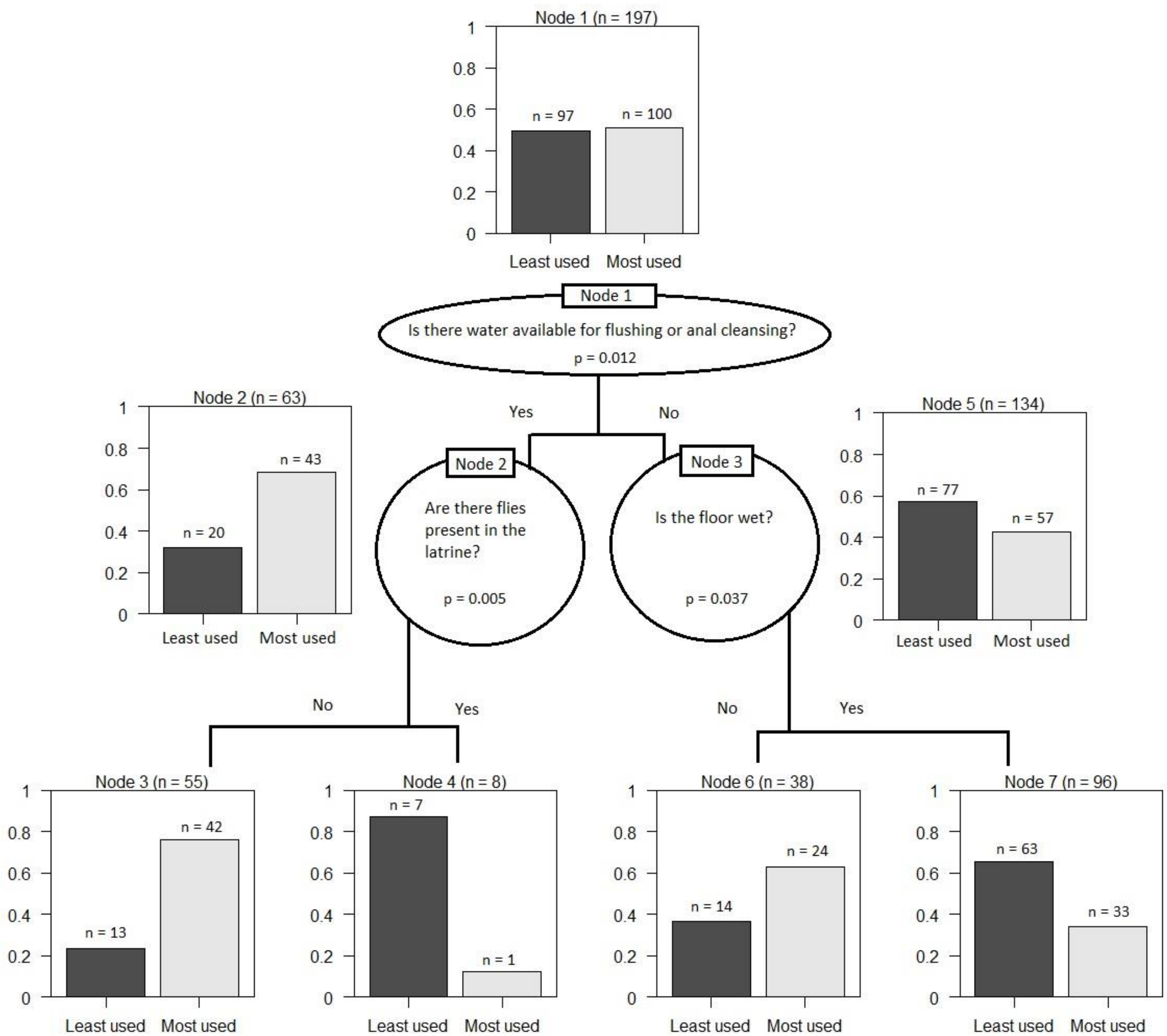


Figure 4. CIT Plot for PLUM recorded use per latrine user categorized using the median cut point (1.25) without Bonferroni correction.

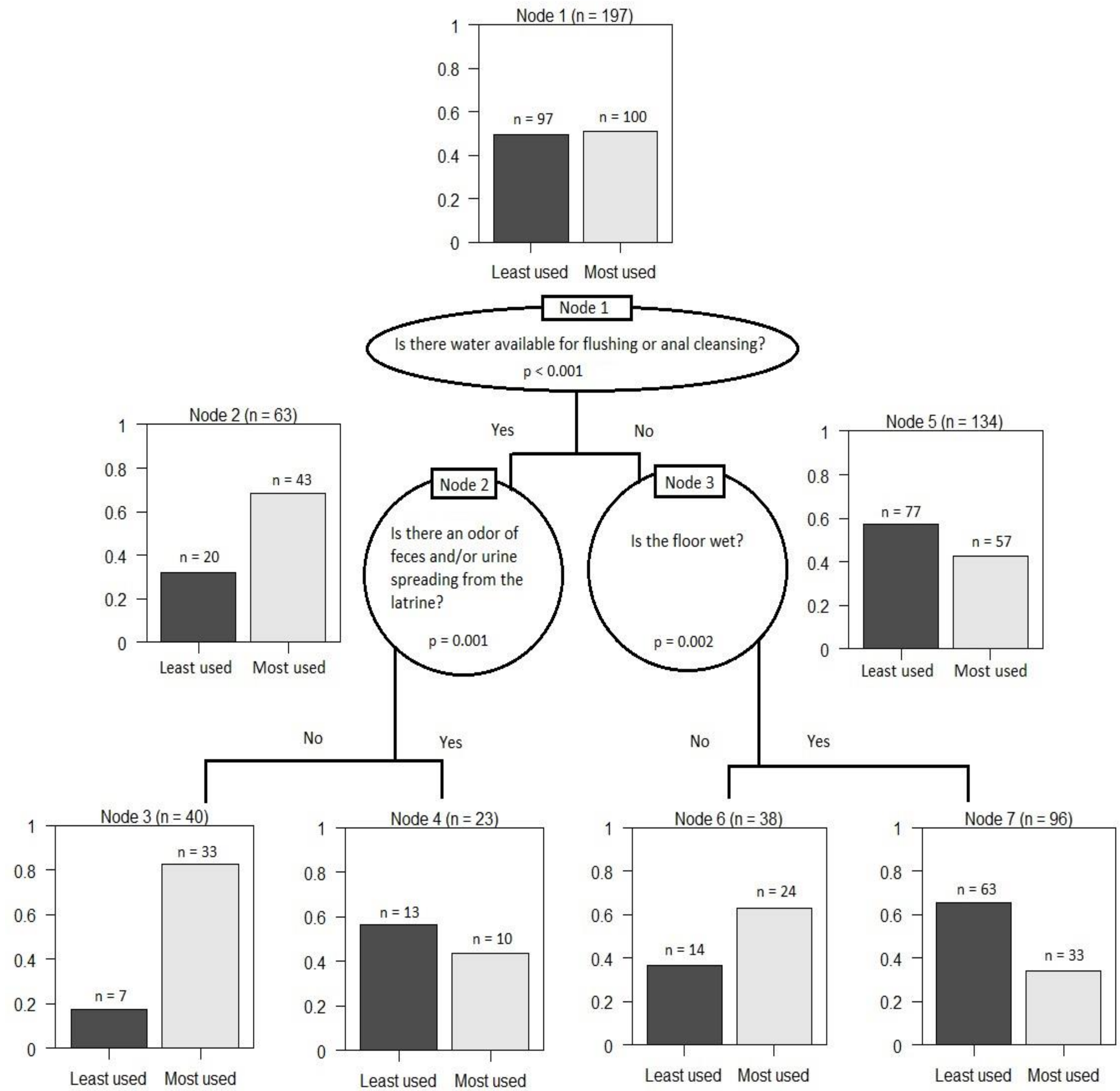


Figure 5. Conditional Inference Tree predicting PLUM recorded use (on a continuous scale) per identified user of this latrine.

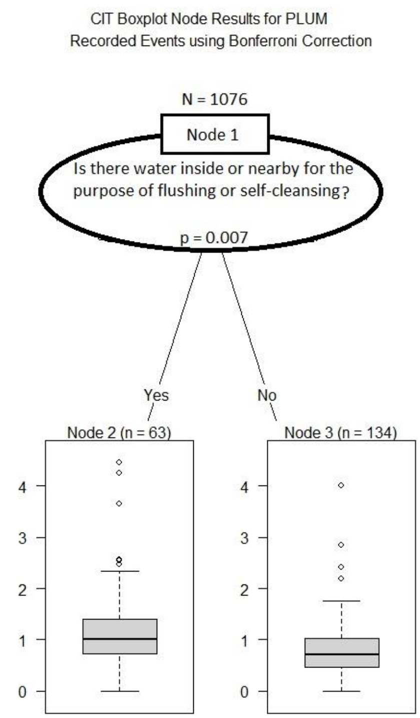


Figure 5a.

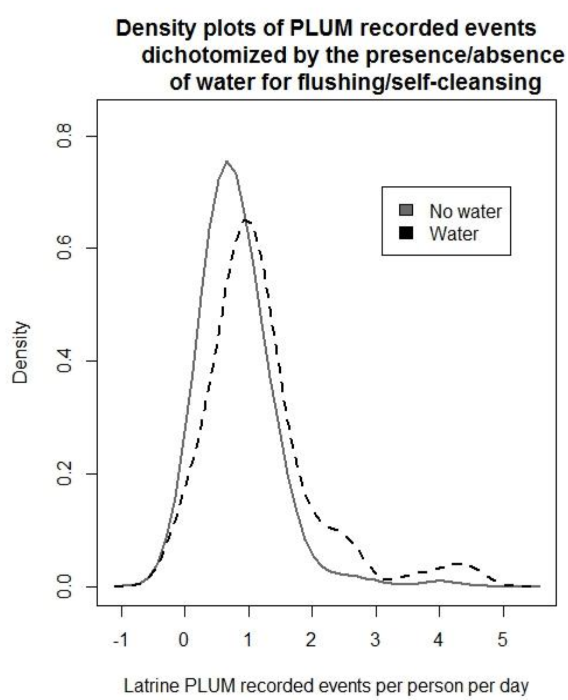


Figure 5b

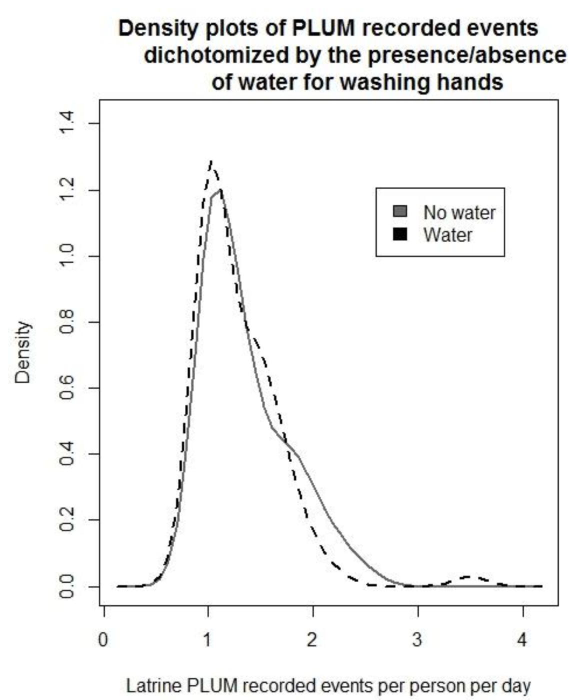


Figure 6. CIT of reported daily latrine usage per identified latrine user without Bonferroni correction.

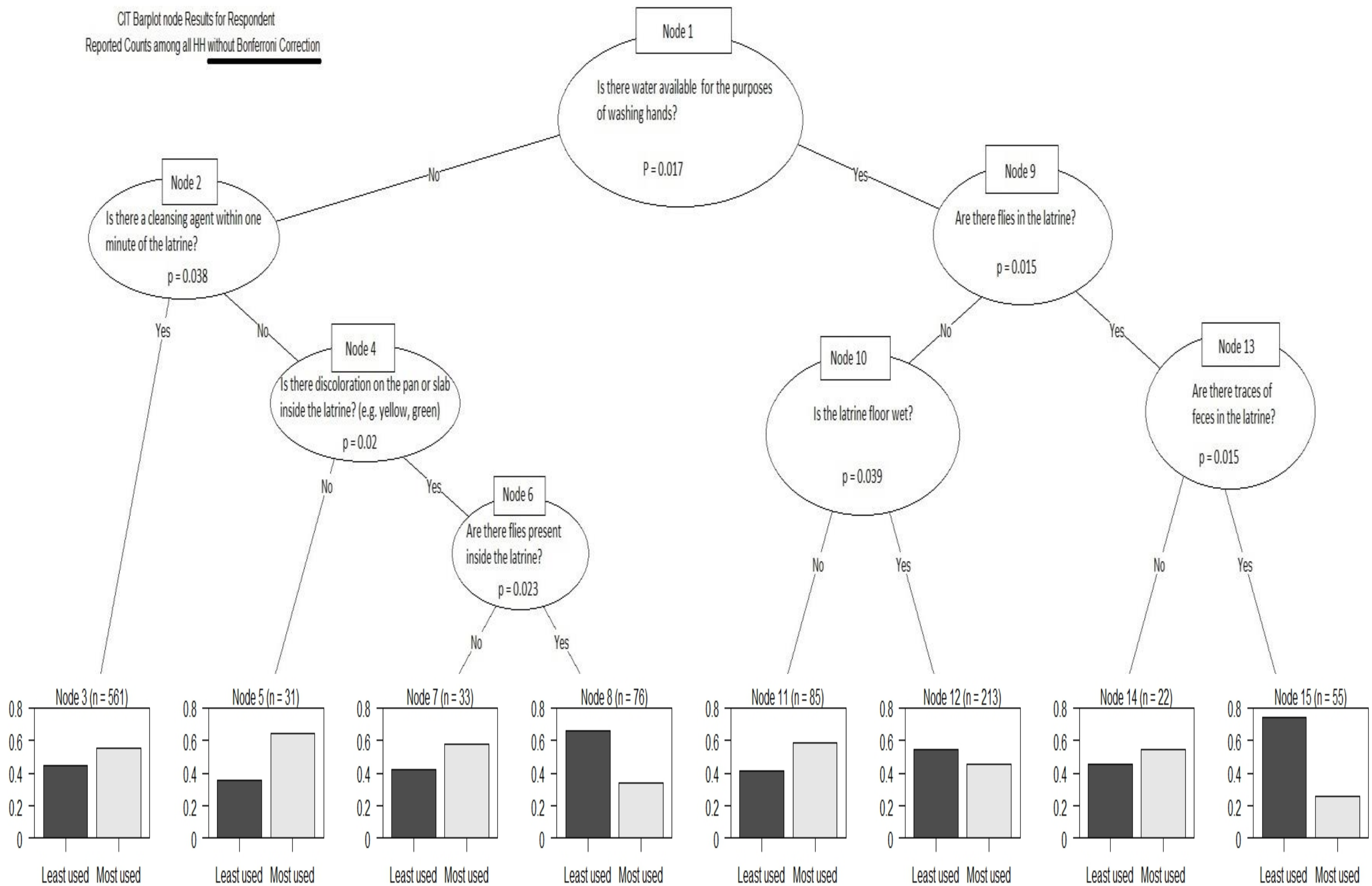


Figure 7. Conditional Inference tree showing reported daily latrine use per identified user categorized as above or below the median.

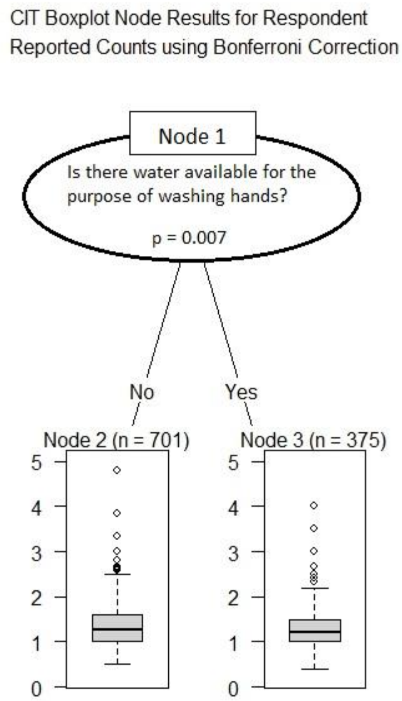


Figure 7a.

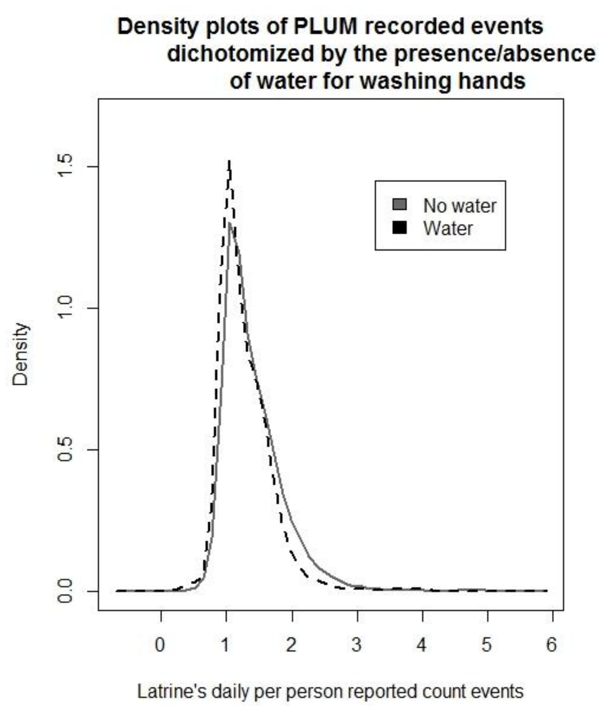
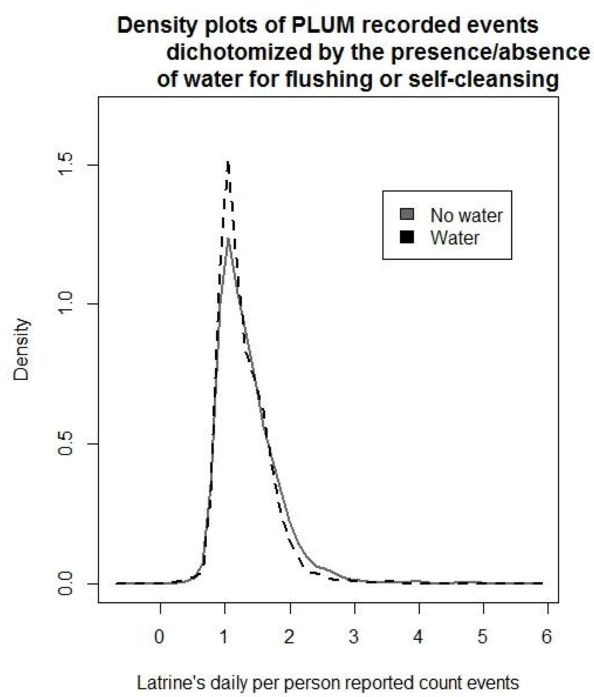


Figure 7b.



REFERENCES

1. World Health Organization. Joint Monitoring Plan Report. 2013,
2. Economic UNDo, Information UNDoP. *The millennium development goals report 2009*. United Nations Publications; 2009.
3. World Health Organization. Progress on sanitation and drinking water - 2015 update and MDG Assessment. WHO Library Cataloguing-in-Publication Data: INICEF and World Health Organization 2015, 2015.
4. Feachem R, Mara DD, Bradley DJ. *Sanitation and disease*. John Wiley & Sons Washington DC, USA:; 1983.
5. Prüss-Üstün A, Bos R, Gore F, et al. *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health*. World Health Organization; 2008.
6. Ziegelbauer K, Speich B, Mausezahl D, et al. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *PLoS medicine* 2012;9(1):e1001162.
7. Clasen T, Boisson S, Routray P, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *The Lancet Global health* 2014;2(11):e645-53.
8. Corrales LF, Izurieta R, Moe CL. Association between intestinal parasitic infections and type of sanitation system in rural El Salvador. *Tropical medicine & international health : TM & IH* 2006;11(12):1821-31.
9. Spears D, Ghosh A, Cumming O. Open defecation and childhood stunting in India: an ecological analysis of new data from 112 districts. *PloS one* 2013;8(9):e73784.
10. Bagcchi S. Open defecation is associated with raised risk of preterm birth among Indian women. *BMJ (Clinical research ed)* 2015;351:h4020.

11. Sahoo KC, Hulland KR, Caruso BA, et al. Sanitation-related psychosocial stress: A grounded theory study of women across the life-course in Odisha, India. *Social science & medicine (1982)* 2015;139:80-9.

12. Haile M, Tadesse Z, Gebreselassie S, et al. The association between latrine use and trachoma: a secondary cohort analysis from a randomized clinical trial. *The American journal of tropical medicine and hygiene* 2013;89(4):717-20.

13. Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bulletin of the World Health Organization* 2003;81(3):197-204.

14. Organization WH. Preventing diarrhoea through better water, sanitation and hygiene: exposures and impacts in low-and middle-income countries. 2014.

15. Assembly UG. Transforming our world: the 2030 Agenda for Sustainable Development. *see, related, The Global Initiative Against Transnational Organized Crime,"Organized Crime: A Cross-Cutting Threat to Sustainable Development"(Geneva: January 2015)* 2015.

16. Kar K, Chambers R, Plan U. *Handbook on community-led total sanitation*. Plan UK London; 2008.

17. Kar K, Pasteur K. Subsidy or self-respect?: Community led total sanitation; an update on recent developments. 2005.

18. Bartram J, Charles K, Evans B, et al. Commentary on community-led total sanitation and human rights: should the right to community-wide health be won at the cost of individual rights? *Journal of Water and Health* 2012;10(4):499-503.

19. Campaign UNM. End Open Defecation Campaign. United Nations Millennium Campaign, 2016.

20. Salient Features of the (Total Sanitation Campaign) Program. Website for Ministry of Rural Development: Indian Ministry of Rural Development, 2010, (Department of Drinking Water and Sanitation (Ministry of Rural Development

21. Patil SR, Arnold BF, Salvatore AL, et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLoS medicine* 2014;11(8):e1001709.

22. Howes M, Islam MA. COMMUNITY-LED TOTAL SANITATION AND ITS SUCCESSORS IN BANGLADESH.

23. Mehta L, Movik S. *Shit matters: The potential of community-led total sanitation*. Practical Action Publishing; 2011.

24. Chambers R. Going to Scale with Community-Led Total Sanitation: Reflections on Experience, Issues and Ways Forward. *IDS Practice Papers* 2009;2009(1):01-50.

25. Ahmed S. Community-Led Total Sanitation in Bangladesh–Chronicles of a People's Movement. Presented at IDS Conference paper2008.

26. Jenkins MW, Freeman MC, Routray P. Measuring the safety of excreta disposal behavior in India with the new Safe San Index: reliability, validity and utility. *International journal of environmental research and public health* 2014;11(8):8319-46.

27. Manun'Ebo M, Cousens S, Haggerty P, et al. Measuring hygiene practices: a comparison of questionnaires with direct observations in rural Zaire. *Tropical medicine & international health : TM & IH* 1997;2(11):1015-21.

28. Clasen T, Fabini D, Boisson S, et al. Making sanitation count: developing and testing a device for assessing latrine use in low-income settings. *Environmental science & technology* 2012;46(6):3295-303.

29. Fisher RJ. Social desirability bias and the validity of indirect questioning. *Journal of consumer research* 1993:303-15.

30. Sinha A, Nagel CL, Thomas E, et al. Assessing Latrine Use in Rural India: A Cross-Sectional Study Comparing Reported Use and Passive Latrine Use Monitors. Yet unpublished, 2016.

31. Thomas EA, Tellez-Sanchez S, Wick C, et al. Behavioral Reactivity Associated With Electronic Monitoring of Environmental Health Interventions-A Cluster Randomized Trial with Water Filters and Cookstoves. *Environmental science & technology* 2016.

32. Srigley JA, Furness CD, Baker GR, et al. Quantification of the Hawthorne effect in hand hygiene compliance monitoring using an electronic monitoring system: a retrospective cohort study. *BMJ quality & safety* 2014;23(12):974-80.

33. Breiman L, Friedman J, Stone CJ, et al. *Classification and regression trees*. CRC press; 1984.

34. Zhang H, Bonney G. Use of classification trees for association studies. *Genetic epidemiology* 2000;19(4):323-32.

35. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the RPART routines. Technical Report 61. URL http://www. mayo. edu/hsr/techrpt/61. pdf, 1997.

36. Apté C, Weiss S. Data mining with decision trees and decision rules. *Future Generation Computer Systems* 1997;13(2–3):197-210.

37. Mackinnon MJ, Glick N. Applications: Data Mining and Knowledge Discovery in Databases–An Overview. *Australian & New Zealand Journal of Statistics* 1999;41(3):255-75.

38. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI magazine* 1996;17(3):37.

39. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *Journal of clinical epidemiology* 2001;54(6):603-9.

40. Hothorn T, Hornik K, van de Wiel MA, et al. A Lego System for Conditional Inference. *The American Statistician* 2006;60(3):257-63.

41. Lemon SC, Roy J, Clark MA, et al. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Annals of behavioral medicine* 2003;26(3):172-81.

42. Rokach L, Maimon O. *Data mining with decision trees: theory and applications*. World scientific; 2014.

43. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics* 2006;15(3):651-74.

44. Rice J. *Mathematical statistics and data analysis*. Nelson Education; 2006.

45. Kim H, Loh W-Y. Classification trees with unbiased multiway splits. *Journal of the American Statistical Association* 2011.

46. Speybroeck N. Classification and regression trees. *International journal of public health* 2012;57(1):243-6.

47. Gass K, Addiss DG, Freeman MC. Exploring the relationship between access to water, sanitation and hygiene and soil-transmitted helminth infection: a demonstration of two recursive partitioning tools. *PLoS neglected tropical diseases* 2014;8(6):e2945.

48. Salis KL, Kliem S, O'Leary KD. Conditional inference trees: a method for predicting intimate partner violence. *Journal of marital and family therapy* 2014;40(4):430-41.

49. Delea M. Report on a Study to Independently Assess Latrine Coverage and Use under BRAC's WASH II Project in Bangladesh. Emory University, 2015.

50. Delea M. You can count on me: Employing limits and indices of agreement to compare respondent-reported and instrument-recorded latrine utilization measures in rural Bangladesh. *Unpublished*, 2016.

51. Delea M. Figure 2. Flow of PLUM data capture. In: background Lumo-Aw, ed. *Microsoft Word*, 2015.

52. Heaton K, Radvan J, Cripps H, et al. Defecation frequency and timing, and stool form in the general population: a prospective study. *Gut* 1992;33(6):818-24.

53. RStudio Team. R Studio: Integrated Development Environment for R. Boston, MA: RStudio, Inc., 2015.

54. Hothorn T, Hornik K, Zeileis A. ctree: Conditional Inference Trees.

55. R Development Core Team. R: A language and environmenet for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2010.

56. Bender R, Lange S. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology* 2001;54(4):343-9.

57. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilit\`{a}. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* 1936;8:3-62.

58. Breiman L. Random forests. *Machine learning* 2001;45(1):5-32.