Development and External Validation of a Logistic Regression Derived Algorithm to Estimate a Twelve-Month Open Defecation-Free Status

Warren Mukelabai Simangolwa[1*]
[1] Independent Health Economist
Lusaka, Zambia
Email: mukewarren@gmail.com
*Corresponding author

Key words: Chiefdom, community-led total sanitation (CLTS), District Health Information Software (DHIS2), Prognostic model, open defecation free (ODF)

1 TARGET AUDIENCE

This paper aimed at all development actors working in, for and with rural communities. It is particularly beneficial to rural households, community-led total sanitation (CLTS) volunteers, toilet masons, traditional leaders, rural water sanitation & hygiene (WASH) practitioners, local and national government officers and national & international non-governmental organisations (NGOs).

2 BACKGROUND

In Zambia, a chiefdom is declared open-defecation free (ODF)when all its households in all villages have an “adequate” toilet. An adequate toilet is one that satisfies the following requirements: 1) contains a smooth cleanable floor, 2) has a superstructure that provides privacy, 3) includes a handwashing station with soap and, 4) has a lid or vent valve to prevent flies. A household toilet that includes all four requirements is considered “adequate”. If not every village in the chiefdom contains adequate toilets, the chiefdom is denoted with an open defecation (OD) status. Maintaining the ODF status in chiefdoms post verification and certification, however, has become increasingly challenging.

To maintain the ODF status, a chiefdom must adopt appropriate interventions that focus on the four adequacy parameters. However, in the absence of cost-effective systematic approaches, the process of identifying the villages in a chiefdom at most risk of not maintaining adequate toilets—reverting back to an OD status—can be costly to both chiefdoms and the Government of Zambia.

3 PURPOSE

After attaining ODF status, utilising non-systematic follow-up interventions in villages can be costly and unsustainable. This paper aims to develop a simple systematic tool to identify villages at high risk of losing ODF status. Equipped with this information, decision-makers can more wisely prioritise and allocate scarce human, financial, logistical and other associated resources for ODF sustainability interventions.

4 METHOD

The study developed a systematic approach to predict when and how an ODF chiefdom in Zambia will revert back to open defecation. The study followed household data collected from 67 villages in the North Zambian Chungu chiefdom, for a period of 12 months. The study used a WASH reporting tool with real time data entered at village-level by community volunteers. The study saw limitations in data capture and reporting, with some of the consulting data set reported as missing. It was assumed that data missingness was a random phenomenon. However, a small fraction of villages had missing data associated to specific variables. Only villages with complete data sets were analysed; a complete case analysis approach was used.

To test whether or not the systematic approach developed could be transferred between chiefdoms, 200 computer-generated samples of fictional chiefdoms (test cases) were created using the Chungu chiefdom data set. This testing of the test cases was conducted using the Harrell method that allowed for improvement of the developed approach.

The improved systematic approach was further tested for its power to predict the transition of a chiefdom from ODF to OD status, using data from a different the Chabula chiefdom of the Northern province.

5 RESULTS

The systematic approach developed showed that a chiefdom can revert from an ODF status back to an OD status with time and with increases in the village population. Furthermore, a chiefdom is more likely to revert to an OD status if there is a higher number of toilets built as a result of community-led total sanitation (CLTS) projects compared to those built prior to any CLTS intervention.

For a model to be acceptable, it must be able to correctly predict the loss of ODF status based on measurable parameters. In addition, it must also correctly distinguish between a successfully maintained ODF status and an ODF reversion to an OD status. The model developed successfully accomplished both criteria.

6 IMPLICATIONS FOR TARGET AUDIENCE

The systematic approach developed in this study uses parameters that are easily accessible to the chiefdom and the Zambian Government, using the community led total sanitation (CLTS) reporting protocol utilised in Zambia. However, cultural, social cohesion, geographical and socio-economic factors were only sparingly considered in this study. Furthermore, the scientific approach applied in the model development phase excluded factors that were not significantly related to ODF status loss. These factors, however, may be of importance to decision-makers and as a consequence may limit the overall outputs of the model. As a consequence of these and other limitations, the results of the study should be applied with caution. In addition, if the results are to be applied in other contexts outside Zambia, uniformity in the definition of the ODF status must be assessed.

Development and External Validation of a Logistic Regression Derived Algorithm to Estimate a Twelve-Month Open Defecation-Free Status

Warren Mukelabai Simangolwa[1*]
[1] Independent Health Economist
Lusaka, Zambia
Email: mukewarren@gmail.com
[*]Corresponding author

Abstract: *Appropriate open defecation free (ODF) sustainability interventions are key to mobilising communities to consume sanitation and hygiene products and services that enhance quality of life and result in embedded behavioural change. This study aims to develop a logistic regression derived risk algorithm to estimate the risk of the loss of ODF status over a 12-month period, and to externally validate the model using an independent data set. ODF status loss occurs when one or more toilet adequacy parameters is no longer present for one or more toilets in a community. Data collected in the Zambia district health information software for water sanitation and hygiene management was utilised in this study. Datasets for the Chungu and Chabula chiefdoms were selected for this study. The data was collected from the date of attainment of ODF status (October 2016) for a period of 12 months until September 2017. The Chungu chiefdom data set was utilised as the development data set whilst the Chabula chiefdom data set was utilised as the validation data set. Data was assumed to be missing at random and the complete case analysis approach was used. The events per variables were satisfactory for both the development and validation data sets. Multivariable regression with a backwards selection procedure was used to decide candidate predictor variables with p values less than 0.05 meriting inclusion. To correct for optimism, the study compared amount of heuristic shrinkage by comparing the model's apparent C-statistic to the C-statistic computed by non-parametric bootstrap resampling. In the resulting model, an increase in the covariates 'months after ODF attainment', 'village population' and 'latrine built after CLTS', were all associated with a higher probability of ODF status loss. Conversely, an increase in the covariate 'presence of a handwashing station with soap', was associated with reduced probability of ODF status loss. The predictive performance of the model was improved by the heuristic shrinkage factor of 0.988. The external validation test confirmed good prediction performance with an area of 0.85 under the receiver operating characteristic curve and no significant lack of fit (Hosmer-Lemeshow test: $p = 0.246$). The results of this study must be interpreted with caution in context where ODF definitions, cultural and other factors are different from those described in the study.*

Key words: Chiefdom, community-led total sanitation (CLTS), District Health Information Software (DHIS2), Prognostic model, open defecation free (ODF)

1 BACKGROUND

Achieving and sustaining open defecation free (ODF) status has increasingly become a shared goal for communities, interventionists, non-governmental and the Government of Zambia. Poor access to sanitation and hygiene infrastructure negatively impacts progress on agreed international targets on health, poverty and human dignity (Roche, et al., 2017; Hutton & Chase, 2016).

In 2015, only 39% of the global population used safely-managed sanitation measures, with two in every five persons living in rural areas. As of 2015, 892 million people worldwide still practice open defecation (World Health Organization , 2017). The diseases associated with poor sanitation practice (including open defecation) account for about 10% of the global burden of disease (Prüss-Üstün, et al., 2008; McGinnis, et al., 2017). These include diarrhoeal diseases, acute respiratory infections, malnutrition and tropical diseases such as helminth and schistosomiasis infection (Van Minh & Hung, 2011; Araujo Navas, et al., 2016). Diarrhoea alone accounted for 19% of the deaths in children under the age of five in Sub-Saharan Africa (Mara, et al., 2010). 88% of these cases are attributable to unsafe water, inadequate sanitation, and poor hygiene (Roche, et al., 2017). The lack of access to safe sanitation cost the global economy USD $222.9 billion in 2015, with associated costs linked to mortality, productivity and healthcare (Lixil, WaterAid Japan & Oxford Economics, 2016). The global economic return on sanitation spending is USD $5.5 per US dollar invested (Hutton, 2012).

The key to mobilising communities to move away from poor sanitation practices, such as open defecation, is the use of appropriate sustainability interventions otherwise termed 'Open Defecation Free' (ODF) measures. One such measure is the adoption of "adequate household toilets". An adequate household toilet is one that satisfies the following design requirements: 1) contains a smooth cleanable floor, 2) has a superstructure that provides privacy, 3) includes a handwashing station with soap and, 4) has a lid or vent valve to prevent flies.

A community, such as a chiefdom, can be designed an Open Defecation Free (ODF) status when all of the household in every village contain an adequate household toilet. If not every village in the chiefdom contains adequate toilets; the chiefdom is designated with an open defecation (OD) status.

Once ODF status is granted through verification and certification, this status must be maintained by the community through ongoing maintenance of existing adequate household toilets as well as the construction of new adequate household toilets. The ODF status is not a permanently held status and can be lost or reverted when one or more of the adequacy parameters are no longer present for one or more toilets in a community (Galan, et al., 2013; Njuguna & Muruka, 2017): this trend is otherwise termed ODF status loss.

It is recognised that maintaining the ODF status is challenging due to uncertainties in social cohesion, d and government prioritisation of sanitation, sustainability of toilet and handwashing technologies, sanitation financing, governance, monitoring and sanitation markets (Bongartz, 2016; Odagiri, et al., 2017). Faced with the challenge of ODF status loss, there is limited scientific evidence to guide a systematic approach on mitigating ODF status loss risk factors.

ODF status loss risk prediction can be useful in guiding and subsequently inform the adoption of cost-effective and timely mitigation measures. (Hendriksen, et al., 2013). A risk score is a standardised metric for the likelihood that a variable of interest will experience an outcome (Royston, et al., 2009). ODF status loss risk prediction modelling is essential in identifying and providing appropriate intervention measures for communities at high risk of ODF status loss.

This paper aims to develop a simple systematic tool to identify villages at high risk of ODF status loss. It develops and externally validates a prognostic model to estimate twelve-month status loss risk for a Zambian chiefdom. Equipped with this information, decision-makers can more wisely prioritise and allocate scarce human, financial, logistical and other associated resources for ODF sustainability interventions.

2 MATERIALS AND METHODS

2.1 Source of Data

In Zambia, for a chiefdom to be designated an ODF status, it must be verified and certified that all the chiefdom villages households include an adequate household toilet (Zimba, et al., 2016). The four design parameters that define an adequate household toilet are the following:

(1) a smooth cleanable floor
(2) a super structure providing privacy
(3) a hand washing station containing soap and

an orifice lid Retrospective longitudinal cohort data was extracted from the district health information software for water sanitation and hygiene management information system (DHIS2 WASH MIS) registry data for both the development and validation data sets.
DHIS2 is a free and open-source framework for management of aggregated health information (Manoj, 2013). The DHIS2 is used for collection, validation, analysis, and presentation of aggregate statistical data, tailored to integrated health management activities (Asangansi, 2012). In Zambia, the DHIS2 serves as the national database platform for the Ministry of Local Government and Housing (MLGH) (Biemba, et al., 2017). The DHIS2 WASH MIS is a mobile surveillance real time monitoring tool that is used for sanitation and hygiene monitoring in Zambia (Markle, et al., 2017).
Data in the DHIS2 WASH MIS is entered via a Java-supported mobile phone by community volunteers (community champions). The volunteers are tasked with providing monthly data monitoring support for village groups averaging 10 villages or more. The community volunteer amalgamates data collected from each village's sanitation action group (SAG). The village SAG is responsible for collecting paper-based household level data on parameters related to toilets. They use the water sanitation and hygiene, sanitation action group (WASH SAG) data collection form. Each community champion visits their assigned villages of supervision, during the period from first to the tenth of each month. They are expected to collect and submit aggregated data from each village's WASH SAG data collection form for the previous month by the tenth day of each month. Before the community champion aggregates data and submits through their mobile phone, they randomly pick three households on the WASH SAG form for a spot verification check. A ward government line agency extension officer, conducts a meeting with a champion to assess the quality of the data prior to submission to the DHIS2WAS MIS platform. The following variables are collected from utilising the WASH SAG data collection form: village name, total number of households in a village, village population, number of toilets before and after Community-Led Total Sanitation (CLTS) approaches were implemented, number of toilets fulfilling the adequate household toilet definition.

2.2 Interventions and Village Selection

The Chungu and Chabula chiefdoms were utilised as the development and validation datasets for this study. Data for both these chiefdoms was extracted from the DHIS2 WASH MIS platform. Chungu and Chabula chiefdoms are two of the five chiefdoms in the Luwingu district of the Northern Province of Zambia.

Sanitation and hygiene interventions were introduced in the two chiefdoms in August 2014 through the local district council with support from its sanitation and hygiene partners. The Community-Led Total Sanitation (CLTS) approach was used for sanitation demand creation in both Chungu and Chabula chiefdoms. CLTS is designed to mobilise individuals in action to eliminate open defecation as a whole community (Harter, et al., 2018). It is a participatory approach in which facilitators visit villages and trigger awareness of sanitation practices and subsequently perform follow-up visits to villages to generate a community-wide effort to become an open defecation free (ODF) status holder (Crocker, et al., 2017).

To build capacity of CLTS facilitators in implementing district-wide activities in Luwingu district, CLTS district and sub-district level implementers ere trained in the CLTS methodology through the District Water Sanitation and Hygiene Education Committee (DWASHE) with the assistance of the CLTS national coaches. The cadres were periodically followed up to review their application of CLTS methodology. The DWASHE is a multi-stakeholder representative body for all government line agencies, donors, local non-governmental organisations (NGOs), civil society organisations and the private sectors involved in the governance and implementation of sanitation interventions at district level (Lungu & Harvey, 2009; Kanyamuna, 2010). To steer quality and effectiveness in implementation targets, strategies, standards, norms and approaches were aligned through the development and use of the Luwingu District Total Sanitation Plan 2014 to 2017 through a multi-stakeholder process. All trained cadres at district and sub-district level were trained in the use of the DHIS2 WASH MIS platform and provided with Java-supported phones, WASH SAG data forms and bicycles to enhance effectiveness in reporting. The baseline data in 2014 revealed that 38% of the population in the Luwingu district had access to toilets whereas 0% had a handwashing station with soap (SNV, 2018).

Chief Chungu and Chabula's chiefdoms were sampled to determine their ODF status on the 20th of October 2016 (Mutyoka & Makombo, 2016; Kachemba, 2016). The Chungu chiefdom is led by her Royal Highness Chieftainess Chungu and is located 58 kilometres from the Luwingu district administration. The Chiefdom has a population of 29,840 people across 220 villages; 34 villages in Kampemba, 42 villages in Kafinsa, 38 villages in Ilambo, 52 villages in Mufili and 54 villages in Mulalashi wards (Central Stattistics Office , 2010). Villages are organised, small local communities of people in daily face-to-face interaction (Drennan & Peterson, 2006). The economy for the Chungu chiefdom is agricultural, with maize as the major cash crop.

The Chabula chiefdom is led by chief Chabula and is located 85 kilometres from the Luwingu district administration. The chiefdom has a population of 14,112 people across 117 villages; 79 villages in the Bwalinde and 33 villages in Ibale wards (Central Stattistics Office , 2010). The Chabula chiefdom's economy is a mix of agriculture and fish farming. The chiefdom borders Lake Bangweulu at Nsombo, a principal town at the northern part of the lake, which supports fish farming (Wikipedia contributors, 2018). Figure 2 shows the location of Chief Chungu and Chief Chabula chiefdoms.

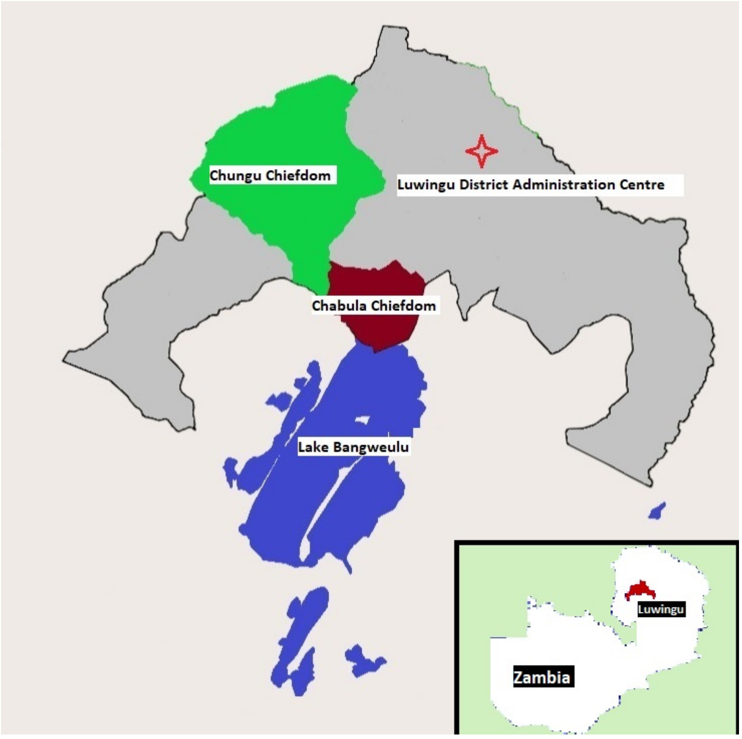

Figure 1: Location of Chungu and Chabula Chiefdom's in Zambia

The village population data collected for the development model includes 67 villages of 220 villages assigned to the Ilambo, and Mufili wards in Chungu chiefdom. In the validation dataset of the Chabula chiefdom, the village population data for 56 of the total 117 villages in the Bwalinde and Ibale wards were used.

To be included for both the development and validation datasets, only the villages that had data imputed for at least 3 of the 12 months of the year (i.e. at least 25%) were considered. There are no agreed estimates in literature on the amount of allowable missing data. Differing views are held by researchers on the appropriate cut-off point for data missingness but there have been suggestions as high as 20% is appropriate (Chao-Ying Joanne, et al., 2006). Table 1 shows the data captured for the study.

Table 1 Village selection for the development and validation models

Chiefdom	**Ward**	**Total number of villages**	**Villages with data**	**Villages without data**	**Villages with at least 3 months of data**
Chungu	Kampemba	34	0	34	0
	Kanfisa	42	7	35	0
	Ilambo	38	37	1	37
	Mufili	52	30	22	30
	Mulalashi	54	15	39	0
Total		220	90	131	67
Chabula	Bwalinde	79	67	12	23
	Ibale	33	33	0	33
Total		117	87	30	56

2.3 Outcome

The measured outcome in both the development and validation models, was ODF status loss. The ODF status loss variable in each village was derived by subtracting the monthly number of adequate toilets from the total monthly number of households in the village. Any non-zero result was assumed as an ODF status loss.

2.4 Predictors

The development and validation datasets were cleansed to ensure that the number of households in a village was always greater than or equal to the number of toilets defined as "adequate". Four cases with this inconsistency were retrieved and evaluated. The study investigated historical cases to ensure the data cleansing minimised the introduction of any new data.

The following predicator variables were extracted for each village: 1) Champion and SAG meeting, 2) latrine in use, 3) latrine built after CLTS, 4) latrine lids and, 5) latrine with smooth and cleanable floor data. Other predicator variables include: 1) latrine privacy, 2) number of latrines with handwashing with soap stations, 3) number of households and, 4) total village population. These predicator variables and co-variates are defined as following:

- *Champion and SAG meeting* - the monthly number of interactions the community volunteer has with SAG.
- *Latrine in use* - the toilet infrastructure available in the village that are used as toilets
- *Latrine built after CLTS* - the number of latrines built after the CLTS intervention in each village.
- *Latrine lid* – ascertains the vector transmission of faecal matter.
- *Latrine with smooth cleanable floor* – ascertains the technologies to enhance use and maintenance of toilet facilities
- *Latrine privacy* - assesses acceptable latrine infrastructure (i.e. latrine walls, door etc.) that ensures dignity for users.

- *Number of latrines with handwashing with soap stations* – quantifies behavioural change through a proxy measure of the availability of a handwashing station with soap within 10 meters from a toilet

2.5 Sample Size

The study did not calculate a formal sample size but used all available data in the development and validation models from the DHIS2 WASH MIS data to maximise the power and generalisability of the results. There are no generally accepted approaches to estimate sample size in the derivation and validation of studies for risk prediction models. Some have suggested having at least 10 events per candidate variable (EPV) for the derivation of a model (Pavlou, et al., 2015). The EPV for both the development and validation models for this study were satisfactory (Austin, et al., 2017; Vittinghoff & McCulloch, 2007) and therefore expected to provide estimates that are robust.

2.6 Missing Data

Limitations to data capture and reporting included the following: limited internet connection, loss of software application (DHIS2 WASH MIS) on mobile phones, failure to access villages during the wet season. As a consequence of these limitations, 12% of the development model data were not captured for random villages in random months and is assumed as missing. 53% of the validation model data was not captured and is assumed as missing.
A data point of a village can be classified 'missing' during one data capture event but then 'measured' again at subsequent data capture events. This results in non-monotonic missing data patterns for villages. In this case, the probability of missing values was not related to the value of the observed responses, and thus the data was assumed to be missing completely at random (MCAR) for both the development and validation models (Kang, 2013). In the development model, however, 3% of the data was missing at random (MAR). In this case, the probability of data missing was dependent on the set of observed responses. It is established that an analysis restricted to study participants with complete datasets can be both biased and inefficient (Spratt, et al., 2010) and this is especially a threat to prognostic studies (Vergouw, et al., 2012). However, it has been shown that multiple imputation methods have offered no statistical advantage over complete case analysis in some assessed scenarios (Mukaka, et al., 2016). This study used the complete case analysis for data evaluation. The reduced dataset assumed the removal of the MCAR cases and was shown to still produce valid estimates for statistical modelling as it represents a randomly drawn sub sample dataset of the original dataset (Bennett, 2009).

2.7 Statistical Analysis

Multivariable regression with a backwards selection procedure was used to decide which of the candidate predictor variables should be included in the final prediction model, with a p-value of less than 0.05 ($p < 0.05$) taken as a conservative indicator of inclusion in the dataset.

In backward selection, all the selected covariates are firstly entered at the same time into the model. Subsequently, the variables with the highest p-values are removed based on the Wald test that allows the calculation of the significance level of a predictor. This step is repeated until there are no variables left with a p-value greater than 0.05 (Steyerberg, et al., 2001).
After the backward stepwise regression was performed, the following variables: 1) month, 2) latrine built after CLTS, 3) latrine with handwashing and, 4) latrine privacy, remained in the development model. The variables: 1) champion and, 2) SAG meeting, 3) latrine in use, 4) latrine lids and, 5) latrine with smooth and cleanable floor were excluded. The variable 'village population' was added to the model due to a p-value of 0.059, very close p-value to $p < 0.05$ criterion. The ability to include variables outside the criteria of $p <0.05$ is supported by literature (Hendriksen, et al., 2013).
Prognostic models in general appear to perform better in datasets used to develop the model than in new datasets. In this case, the regression coefficients and the measures of performance for the model are optimistic (Steyerberg, et al., 2004). To correct for the optimism, the study's development model parameter estimates were penalised by a shrinkage factor. To estimate the shrinkage in the model, the study used the Van Howelingen and le Cessie shrinkage estimator (Van Houwelingen, 2001).
In the Van Howelingen and le Cessie shrinkage technique, the regression coefficients are multiplied by the heuristic shrinkage factor and the intercept is re-estimated (Pajouheshnia, et al., 2016). The study's internal validation associated the amount of shrinkage by comparing the model's apparent C-statistic to the C-statistic computed by nonparametric bootstrap resampling. The bootstrap-resampled C-statistic corrects for regression towards the mean and overfitting (Steyerberg, et al., 2001). The internal bootstrap validation C-statistic for the study used the Harrell's nine steps on 200 bootstrap resamples of 688 cases each.
To record overall predictive performance in addition to discrimination, calibration was used. Discrimination is the ability of the risk score to differentiate between villages which do and do not experience an adverse event during the study period (Steyerberg, et al., 2010). This measure is quantified by calculating the area under the receiver operating characteristic (ROC) curve. A value of 0.5 represents chance, $0.7 \leq ROC <0.8$ represents acceptable discrimination, $0.8 \leq ROC <0.9$ represents excellent discrimination, and a value of 1 represents perfect discrimination (Hosmer & Lemeshow, 2000).
Calibration is the ability to accurately assign the correct event probability at all levels of predicted risk (Crowson, et al., 2013). It measures the agreement between the observed and predicted risks. It is computed as the difference between the mean observed risk and the mean predicted risk. To measure calibration, the study used the Hosmer-Lemeshow goodness-of-fit test. In the Hosmer-Lemeshow goodness-of-fit test, model outputs are sorted into equally-sized groups, where the probabilities and true states in these groups are then checked by a χ^2 goodness-of-fit test (Dreiseitl & Osl, 2012).
In each group, the expected number of events for ODF status loss in each group of villages was calculated as the sum of the predicted probabilities for the villages in that group. Whereas the observed number of events for ODF status loss was calculated as the sum of the number of events observed in that group (Crowson, et al., 2013).
The study used the December 2016 revised version of Stata/IC 14.2 for windows to all its statistical analyses (StataCorp, 2015). The study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement in its reporting (Collins, et al., 2016).

3 RESULTS AND DISCUSSION

3.1 Study Baseline Data

A total of 803 cases were available for the development model and 676 cases for the validation data set. The villages that did not have any data reported over some months in the DHIS2 WASH MIS (i.e. 97 cases), were removed. Table 2 provides a summary of key study characteristics. The final development model had five covariates, with an events per candidate (EPV) equal to 14. The validation model had an EPV of greater than 13.

Table 2 Development and validation models key study characteristics

Characteristic	**Development model: Chungu chiefdom (n = 688)**	**Validation model: Chabula Chiefdom (n = 313)**
Data collection period	October 2016 to September 2017	October 2016 to September 2017
Study design	Retrospective prognostic study	Retrospective prognostic study
Setting	Chief Chungu ODF chiefdom predominately agricultural economy	Chief Chabula ODF Chiefdom predominantly fish farming economy
Inclusion criteria	All villages reported in the DHIS2 WASH MIS with two or more months' worth of data in the system	All villages reported in the DHIS2 with two or more months' worth of data in the system
Outcome	ODF or ODF status loss	ODF or ODF status loss
Total number of villages included in study	67	56
ODF Status	688	313
ODF	618 (89.8%)	246 (78.59%)
ODF status loss	70 (10.2%)	67 (21.41%)

3.2 Internal Validation

The ODF status loss risk model, when internally validated, had a high discrimination (apparent C-statistic = 0.8114, 95% CI = 0.75049 to 0.87240). A Hosmer-Lemeshow of 0.2377 confirmed no significant difference between observed and predicted ODF status loss. The mean for the observed risk seems to be accurately estimated by the mean for the predicted risk grouped by tenths of predicted risk.

3.3 Bootstrap Validation

The difference in the true probabilities from the model's prediction was at 95.9% and the percentage of fit due to noise was 4.12%. This overfitting was estimated by the heuristic shrinkage estimator. After bootstrapping with 200 resamples, the optimism-corrected C-statistic was 0.802. The predicted equation for the ODF status loss model was penalised to account for 1.2% overfitting using the Harrell method (Harrell , et al., 1996) with a heuristic shrinkage factor of 0.988.

3.4 External Validation

The model from the development cohort was penalised for overfitting and applied to an external validation set. The discrimination of the model in the new data set was high (C-statistic=0.844, 95% CI 0.788 to 0.899) and the model was well calibrated with a Hosmer-Lemeshow of p = 0.246 confirming no significant difference between observed and predicted ODF status loss.
The odds ratio of the final model in the prediction equation (Equation 1) indicates that for each one-month increase when other covariates are held constant, the odds of ODF status loss increases by 27.7%, whereas a one person increase in the population for a community with all covariates held constant increases the risk of ODF status loss by 1%. Furthermore, for each latrine that is constructed as a result of a CLTS intervention, the odds of ODF status loss increase by 11.8% when all covariates are held constant. For each handwashing station with soap for a household, the odds of a community experiencing ODF status loss reduce by 20.2% whereas the odds increase by a factor of 4.7 for each increase in a toilet facility that has privacy, all things being equal.

$$P = \frac{1}{1+e^{-x}} \tag{1}$$

where;
x = -4.459 (0.245t + 0.01n + 0.112ℓ - 1.599h + 1.540v)
P = ODF status loss
t = time (months)
n = village population (people)
ℓ = number of latrines built after CLTS (latrines)
h = latrines with handwashing with soap facility (handwashing facility)
v = latrine privacy (latrine wall and door or suitable acceptable substitutes)

The above example results agree with literature accounts on the influence of variables such as: time after ODF attainment, population growth, CLTS interventions and toilet quality on the sustainability of ODF maintenance for villages. ODF sustainability evaluations have estimated an annual ODF status loss rate of 10% per year; with a five-year status loss rate of up to 50% (Thomas, 2016; Tyndale-Biscoe, et al., 2013). ODF status loss in these cases refer to a return in open defecation.

The provision of technical support (ODF sustainability measures) by local and external support agencies following ODF status attainment has been identified as a significant factor in ODF sustainability (Tyndale-Biscoe, et al., 2013). These ODF sustainability measures are collaborative interactions of streamlined government line agencies, local natural leaders and chiefs (Balfour & Singh, 2015). There is also evidence suggesting that the ODF status of a community can still be sustained despite limited follow-up enforcement following ODF status obtainment (Thomas, 2016).
This study also reinforces the findings from a study in Ghana and Ethiopia on sustainability of CLTS outcomes. The study asserts that CLTS is not an appropriate intervention in cases where the baseline toilet coverage is low and local toilet technologies are poor (Crocker, et al., 2017). Poor quality latrines can cause households to revert back to open defecation (Tyndale-Biscoe, et al., 2013; Mosler, et al., 2018). The CLTS methodology encourages communities to construct low cost, simple toilets that leverage on locally available materials (Mosler, et al., 2018). The implementation of CLTS is premised on the assumption that households will upgrade the initial simple low-cost toilet hardware to increasingly higher standards toilets (Khale & Ashok , 2008). Undeveloped sanitation supply chains and poor sanitation markets coupled by unstable soil conditions however, contribute to ODF status loss (Munkhondia, et al., 2018; Garn, et al., 2017). Poor sanitation markets further exacerbate ODF status loss due to their influence on the access to handwashing products such as soap. A correlation in lack of handwashing with soap and ODF status loss was established at 8% (Shivanarain & Nancy, 2015). A cross sectional study to ascertain the association of ODF status loss and the strength of social norms in Indonesia for 587 households after a two-year ODF period estimated the status loss rate at 14.5% (Odagiri, et al., 2017).

4 STRENGTHS

This study's strength is an exhaustive use of the easily accessible DHIS2 WASH MIS repository data. Furthermore, the external validation of the chiefdom's nomadic fish farming and peri-urban social economic setup in selected parts of communities, were distinct characteristics to that of the development model. The study further aligned its statistical plan and results to the TRIPOD statement to ensure for quality and standards. To ensure for transparency, the study used two vigorous and robust measures to correct the overfitting in the model; the heuristic shrinkage estimator and the Harrell method for bootstrapping. The final model was then subject to correction for optimism.

5 LIMITATIONS

Bias of misrepresentation of data by the community cadres cannot be completely ruled out. Whilst the adverse events satisfied recommendations in literature of EPV of greater than 10, adverse events were small in the validation cohort than what is being advocated for in recent literature of at least a 100 adverse events (Collins, et al., 2016). Furthermore, recent literature has advocated for larger EPV values of between 20 and 50 (Austin & Steyerberg, 2014). Complete case analysis was undertaken in the presence of 3% missing at random (MAR) data. The study had a strict adherence to statistical rigor in the selection of study covariates. The relationship between covariates was not explored in this study.

The results can be considered when applying future interventions and the prioritisation aspects of service provision. However, cultural, geographical, socio-economical and other factors may have a particular impact on particular predicators and should be considered in all applications of these results. Caution should be exercised when interpreting the results in contexts where the ODF definitions differ from those defined in this paper.

6 CONCLUSION AND IMPLICATIONS

The study has developed and externally validated a novel population risk prediction algorithm that can predict a twelve-month ODF status loss risk for communities with multiple risk factors. The study utilised monthly available data collected through the DHIS2 WASH MIS platform. This prognostic tool represents a novel and yet simple approach to assessing the risk of ODF status loss that can be used to inform prioritisation of interventions by the following groups and individuals: sanitation action groups at village level, community champions and government extension officers, district officers, provisional officers and national level and general implementing organisations.

Future research should focus on using prospective data to develop and externally validate the ODF status loss prognostic tool in a larger EPV sample (e.g. EPV >20). Furthermore, a controlled qualitative study should be conducted to ascertain factors that explain the negative influence on ODF status loss post ODF status obtainment due to the following variables village population, quality of toilet infrastructure after CLTS and toilet privacy technologies.

7 OTHER INFORMATION

7.1 Supplementary Information

The web calculator for the risk algorithm model is accessible through the following link.

7.2 Funding

No funding was obtained to undertake the study.

7.3 Ethical Considerations

Ethical approval was not necessary

8 REFERENCES

Araujo Navas, AL, Hamm, NA, Soares Magalhães, RJ & Stein, A 2016, ‘Mapping Soil Transmitted Helminths and Schistosomiasis under Uncertainty: A Systematic Review and Critical Appraisal of Evidence’, *PLoS Neglected Tropical Diseases* , vol. 10, no. 2, doi:10.1371/journal.pntd.0005208.

Asangansi, I 2012, ‘Understanding HMIS Implementation in a Developing Country Ministry of Health Context - an Institutional Logics Perspective’, *Online Journal of Public Health Informatics* , vol. 4 no. 3, http://doi.org/10.5210/ojphi.v4i3.4302.

Austin, PC & Steyerberg, EW 2014, 'Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models', *Statistical methods in medical research,* vol. 26, no. 2, pp. 796-808, doi:10.1177/0962280214558972.

Austin, PC, Allignol, A & Fine, JP 2017, 'The number of primary events per variable affects estimation of the subdistribution hazard competing risks model', *Journal of Clinical Epidemiology ,* vol. 83, pp. 75-84. https://doi.org/10.1016/j.jclinepi.2016.11.017.

Balfour, N & Singh, S 2015, *Sustainability of ODF Practices in Kenya,* viewed March 7 2018, https://www.unicef.org/esaro/UNICEF-FN-ODF-Sustainability.pdf

Bennett, DA 2009, 'How can I deal with missing data in my study?', *Australian and New Zealand Journal of Public Health , vol. 25 no.* 5, pp. 464-469, https://doi.org/10.1111/j.1467-842X.2001.tb00294.x.

Biemba, G, Chiluba, B, Yeboah-Antwi, K, Silavwe, V, Lunze, K, Mwale, RK, Russpatrick, S & Hamer, DH 2017, 'Mobile-Based Community Health Management Information System for Community Health Workers and Their Supervisors in 2 Districts of Zambia', *Global Health: Science and Practice,* vol. 5, no. 3, pp. 486–494, http://doi.org/10.9745/GHSP-D-16-00275.

Bodner, TE 2008, 'What Improves with Increased Missing Data Imputations?', *A Multidisciplinary Journal ,* vol. 15, no. 4, pp. 651-675, https://doi.org/10.1080/10705510802339072.

Bongartz, NV 2016, 'Going beyond open defecation free' in NVP Bongartz (ed.), *Sustainable Sanitation for All: Experiences, Challenges and Innovations*, pp. 1-3, https://doi.org/10.3362/9781780449272.001). Rugby: Practical Action.

Central Statistics Office 2010, *2010 census of population and housing,* viewed December 19, 2017, https://www.zamstats.gov.zm/phocadownload/2010_Census/2010%20Census%20of%20Population%20National%20Analytical%20Report.pdf

Chao-Ying Joanne, P, Harwell, M, Liou, SM & Ehman, LH 2006, 'Advances in missing data methods and implications for educational research' in IS (ed.), *Real data analysis*, pg. 3178, https://pdfs.semanticscholar.org/c557/71f7145072c84a21eefd6f0a04e65f61cd08.pdf). Greenwich: Information Age.

Collins, GS, Ogundimu, EO & Altman, DG 2016, 'Sample size considerations for the external validation of a multivariable prognostic model: a resampling study', *Statistics in medicine,* vol. 35, no. 2, pp. 214-226. doi:10.1002/sim.6787.

Collins, GS, Reitsma, JB, Altman, DG & Moons, KG 2015, 'Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement', *BMC Medicine*, vol. 13, no. 1, doi:10.1186/s12916-014-0241-z.

Crocker, J, Saywell, D & Bartr, J. 2017, ‘Sustainability of community-led total sanitation outcomes: Evidence from Ethiopia and Ghana’, *International Journal of Hygiene and Environmental Health ,* vol. 220, no. 3, pp. 551–557. doi: 10.1016/j.ijheh.2017.02.011.

Crocker, J, Saywell, D, Shields, KF, Kolsky, P & Bartram, J 2017, ‘The true costs of participatory sanitation: Evidence from community-led total sanitation studies in Ghana and Ethiopia’, *The Science of the Total Environment,* vol. 601, no. 602, pp. 1075-1083, http://doi.org/10.1016/j.scitotenv.2017.05.279.

Crowson, CS, Atkinson, EJ & Therneau, TM 2013, ‘Assessing Calibration of Prognostic Risk Scores’, *Statistical methods in medical research ,* vol. 24, no. 4, pp. 1692-706. doi: 10.1177/0962280213497434.

Crowson, CS, Atkinson, EJ & Therneau, TM 2013, ‘Assessing Calibration of Prognostic Risk Scores’, *Statistical methods in medical research ,* vol. 25, no. 4, pp. 1692-706. doi: 10.1177/0962280213497434.

Dreiseitl, S & Osl, M 2012 ‘Testing the Calibration of Classification Models from First Principles, *AMIA Annual Symposium proceedings, AMISymposium,* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540450/pdf/amia_2012_symp_0164.pdf

Drennan, RD & Peterson, CE 2006, ‘Patterned variation in prehistoric chiefdoms’, *Proceedings of the National Academy of Sciences of the United States of America, vol. 103*, no. 11, pp. 3960–3967, http://doi.org/10.1073/pnas.0510862103

Galan, DI, Kim, SS & Graham, JP 2013, ‘Exploring changes in open defecation prevalence in sub-Saharan Africa based on national level indices’, *BMC Public Health ,*vol. 13*, pg.* 527, doi:10.1186/1471-2458-13-527.

Garn, JV, Sclar, GD, Freeman, MC, Penakalapati, G, Alexander, KT, Brooks, P, et al. 2017, ‘The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis’, *International journal of hygiene and environmental health ,* vol. 220, no. 2 Pt B, pp. 329-340. doi:10.1016/j.ijheh.2016.10.001.

Harrell , FE, Lee, KL & Mark, DB 1996, Tutorial in Biostatistics: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, *Statistics in Medicine,* vol. 15, no. 4, pp. 361-387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Harter, M, Mosch, S & Mosler, HJ 2018, How does Community-Led Total Sanitation (CLTS) affect latrine ownership? A quantitative case study from Mozambique. *BMC Public Health,* vol. 18, no. 387, doi:10.1186/s12889-018-5287-y.

Hendriksen, J, GJ, G, De, G & Moons, KG 2013, ‘Diagnostic and prognostic prediction models’, *Journal of Thrombosis and Haemostasis ,* vol. 11, no. s1, pp. 129–41. https://doi.org/10.1111/jth.12262.

Hosmer, DW & Lemeshow, S 2000, *Applied Logistic Regression,* 2nd edn, John Wiley and Sons Inc, New York

Hutton, G 2012, 'Global costs and benefits of drinking-water supply and sanitation interventions to reach the MDG target and universal coverage', *Journal of Water and Health,* vol. 11, no. 1, pp. 1-12. doi:10.2166/wh.2012.105.

Hutton, G & Chase, C 2016, 'The Knowledge Base for Achieving the Sustainable Development Goal Targets on Water Supply, Sanitation and Hygiene', *International Journal of Environmental Research and Public Health*, vol. 13, no. 6, pg. 536. doi: 10.3390/ijerph13060536.

Jinks, R 2012, 'Sample Size for Multivariable Prognostic Models', PhD thesis, University College London, viewed March 3 2018, http://discovery.ucl.ac.uk/1354112/1/PhD_master_doc%20to%20be%20submitted.pdf

Kachemba, N 2016, 'ODF status elates Chabula chiefdom', *The Zambia Daily Mail*, November 29, http://www.daily-mail.co.zm/odf-status-elates-chabula-chiefdom/

Kang, H 2013, 'The prevention and handling of the missing data', *Korean Journal of anesthesilogy, vol. 64, no.* 5, pp. 402–406, doi: 10.4097/kjae.2013.64.5.402.

Kanyamuna BM 2010, 'The Impacct of Implimentating the D-WASHE Programmes in Chanyaya Community-Kafue District, Zambia: What role has National Water Policy (1994), played?', University of Zambia, viewed April 18 2018, http://dspace.unza.zm/handle/123456789/1081

Khale, M & Ashok , D 2008, *The impact of rural sanitation on water quality and waterborne diseases,* viewed June 17 2018, http://www.communityledtotalsanitation.org/sites/communityledtotalsanitation.org/files/IHMP_Impact%20of%20Rural%20Sanitation.doc

Lixil, WaterAid Japan & Oxford Economics 2016, *The true cost of poor sanitation,* viewed April 3 2018, https://www.lixil.com/en/sustainability/pdf/the_true_cost_of_poor_sanitation_e.pdf

Lungu, C & Harvey, P 2009, 'Multi-sectoral decentralized water and sanitation provision in Zambia: rhetoric and reality', in Shaw, RJ (ed.), *Water, sanitation and hygiene - Sustainable development and multisectoral approaches: Proceedings of the 34th WEDC International Conference*, Addis Ababa, Ethiopia, viewed March 11 2018, https://dspace.lboro.ac.uk/2134/28808

Manoj, SM 2013, 'Customising DHIS2 for Maternal and Child Health Information Management in Sri Lanka', *Sri Lanka Journal of Bio-Medical Informatics, vol. 3, no.* 2, pp. 47-54, doi: http://doi.org/10.4038/sljbmi.v3i2.2496.

Mara, D, Lane, J, Scott, B & Trouba, D 2010, ‘Sanitation and Health’ *PLoS Medicine,* vol. 7, no. 11, doi:10.1371/journal.pmed.1000363.

Markle, L, Maganani, A, Katooka, O, Tiwari, A, Osbert, N, Larsen, DA, Winters B 2017, ‘A Mobile Platform Enables Unprecedented Sanitation Uptake in Zambia’, *PLoS Neglected Tropical Diseases,* vol. 11, no. 1, https://doi.org/10.1371/journal.pntd.0005131

McGinnis, S. M, McKeon, T, Desai, R, Ejelonu, A, Laskowski, S & Murphy, HM 2017, ‘A Systematic Review: Costing and Financing of Water, Sanitation, and Hygiene (WASH) in Schools’, *International Journal of Environmental Research and Public Health*, vol. 14, no. 4, pg. 442. doi:10.3390/ijerph14040442.

Mosler, HJ, Mosch, S & Harter, M 2018, ‘Is Community-Led Total Sanitation connected to the rebuilding of latrines? Quantitative evidence from Mozambique’, *PLoS One, vol. 13*, no. 5, doi:10.1371/journal.pone.0197483.

Mukaka, M, White, SA, Terlouw, DJ, Mwapasa, V, Kalilani-Phiri, L & Faragher, BE 2016, ‘Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing?’, *Trials,* vol. 17, no. 341, doi: 10.1186/s13063-016-1473-3.

Munkhondia, T, Simangolwa, WM & Maceda, AZ 2018, ‘CLTS and sanitation marketing: aspects to consider for a better integrated approach’, in NV Petra Bongartz (ed.), *Sustainable Sanitation for All Experiences, challenges, and innovations*, pp. 100, 101, Rugby, UK, Practical Action Publishing Ltd., http://dx.doi.org/10.3362/9781780449272)

Mutyoka, M & Makombo, Y 2016, November). *Chabula and Chungu Chiefdom mass verification and certification report in Luwingu district,* viewed April 11 2018, http://41.77.4.165:6510/www.communityledtotalsanitation.org/sites/communityledtotalsanitation.org/files/LUWINGU%20MVC%20CC%20FINAL%20REPORT%20OCTOBER%202016%20.pdf

Njuguna, J & Muruka, C 2017, ‘Open Defecation in Newly Created Kenyan Counties: A Situational Analysis’, *Journal of health care for the poor and underserved*, vol. 28, no. 1, pp. 71-78, doi: 10.1353/hpu.2017.0009.

Odagiri, M, Muhammad, Z, Cronin, AA, Gnilo, ME, Mardikanto, AK, Umam, K, Asamou, YT 2017, ‘Enabling Factors for Sustaining Open Defecation-Free Communities in Rural Indonesia: A Cross-Sectional Study’, *International journal of environmental research and public health,* vol. 14, no. 1572, pp. 15-18, doi: 10.3390/ijerph14121572.

Pajouheshnia, R, Pestman, WR, Teerenstra, S & Groenwold, RH 2016, ‘A computational approach to compare regression modelling strategies in prediction research’, *BMC Medical Research Method ,* vol. 16, no. 1, pg. 107, doi: 10.1186/s12874-016-0209-0.

Pavlou, M, Ambler, G, Seaman, SR, Guttmann, O, Elliott, P, King, M, Omar RZ 2015, 'How to develop a more accurate risk prediction model when there are few events' *BMJ ,* vol. 351, doi: 10.1136/bmj.h3868.

Prüss-Üstün, A, Bos, R, Gore, F & Bartram, J 2008, *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health*, World Health Organisation, viewed April 14 2018, http://41.77.4.165:6510/whqlibdoc.who.int/publications/2008/9789241596435_eng.pdf

Riley, RD, Hayden, JA, Steyerberg, EW, Moons, KG, Abrams, K, Kyzas, PA, Malats, N, Briggs, A, Schroter, S, Altman DG, Hemingway, H, 2013 'Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research', *PLoS Medicine,* vol. 10, no. 2, https://doi.org/10.1371/journal.pmed.1001380

Roche, R, Bain, R & Cumming, O 2017, 'A long way to go – Estimates of combined water, sanitation and hygiene coverage for 25 sub-Saharan African countries', *PLoS One , vol. 12, no.* 2, doi:10.1371/journal.pone.0171783.

Royston, P, Moons, KG, Altman, DG & Vergouwe, Y 2009, 'Prognosis and prognostic research: developing a prognostic model', *BMJ ,* vol. 338, https://doi.org/10.1136/bmj.b604.

Sharmani , B, Routray, P, Majorin, F, Peletz, R, Boisson, S, Sinha, A, Clasen, T 2013, 'Impact of Indian Total Sanitation Campaign on Latrine Coverage and Use: A Cross-Sectional Study in Orissa Three Years following Programme Implementation', *PLoS ONE, vol. 8, no.* 8, doi:10.1371/journal.pone.0071438.

Shivanarain, S & Nancy, B 2015, *Sustainability of ODF Practices in Kenya,* viewed March 12 2018, https://www.unicef.org/esaro/UNICEF-FN-ODF-Sustainability.pdf

Sinha, A, Nagel, CL, Schmidt, WP, Torondel, B, Boisson, S, Routray, P, Clasen, T 2017, 'Assessing patterns and determinants of latrine use in rural settings: A longitudinal study in Odisha, India', *International Journal of Hygiene and Environmental Health, vol. 220, no.* 5), pp. 906–915, doi: 10.1016/j.ijheh.2017.05.004.

SNV 2018, *Zambia – SSH4A Results Programme [Practice Brief],* viewed April 12 2018, http://41.77.4.164:6510/www.snv.org/public/cms/sites/default/files/explore/download/zambia-ssh4a-rp-epb-final-w.pdf

Spratt, M, Carpenter, J, Sterne, JA, Carlin, JB, Heron, J, Henderson, J, Tilling, K 2010, 'Strategies for Multiple Imputation in Longitudinal Studies', *American Journal of Epidemiology*, vol. 172, no. 4, doi: 10.1093/aje/kwq137.

StataCorp 2015, *Stata Statistical Software* (Release 14 ed.), College Station: TX: StataCorp LP.

Steyerberg, EW 2009, 'Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating: By Ewout W. Steyerberg', *American Journal of Epidemiology,* vol. 170, no. 4, pg. 528, https://doi.org/10.1093/aje/kwp129.

Steyerberg, EW, Borsboom, GJ, van Houwelingen, JC, Eijkemans, MJ & Habbema, DF 2004, 'Validation and updating of predictive logistic regression models: a study on sample size and shrinkage', *Statistics in Medicine,* vol. 23, no. 16, pp. 2567 - 2586. doi: 10.1002/sim.1844.

Steyerberg, EW, Eijkemans, MJ, Harrell, FE & Habbema, DJ 2001, 'Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets', *Medical Decision Making,* vol. 21, no. 1, pp. 45 – 56, doi: https://doi.org/10.1177/0272989X0102100106.

Steyerberg, EW, Vickers, AJ, Cook, NR, Gerds, T, Gonen, M, Obuchowski, N, Pencina, MJ, Kattan, MW 2010, 'Assessing the performance of prediction models: a framework for some traditional and novel measures', *Epidemiology,* vol. 21, no. 1, pp. 128-138, doi:10.1097/EDE.0b013e3181c30fb2.

Thomas, A. 2016, 'Strengthening post-ODF programming:reviewing lessons from sub-Saharan Africa', in P. Bongartz, N. Vernon & J. Fox (eds.), *Sustainable Sanitation for all experiences, challenges, and innovations*, Rugby, UK: Practical Action Publishing, pg. 84, http://dx.doi.org/10.3362/9781780449272

Tyndale-Biscoe, P, Bond , M & Kidd, R 2013, *ODF Sustainability Study,* viewed March 11 2018, http://41.77.4.164:6510/www.communityledtotalsanitation.org/sites/communityledtotalsanitation.org/files/Plan_International_ODF_Sustainability_Study.pdf

UCLA 2018, 'Multiple Imputation in Stata', *Institute for Digital Research and Education*, viewed March 13 2018, https://stats.idre.ucla.edu/stata/seminars/mi_in_stata_pt1_new/.

Van Houwelingen, HC 2001, 'Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy', *Statistica Neerlandica*, vol. 55, no. 1, pp. 17-34. https://doi.org/10.1111/1467-9574.00154.

Van Minh, H & Hung, NV 2011, 'Economic Aspects of Sanitation in Developing Countries', *Environmental Health Insights,* vol. 5, pp. 63–70, http://doi.org/10.4137/EHI.S8199.

Vergouw, D, Heymans, MW, van der Windt, DA, Foster, NE, Dunn, KM, van der Horst, HE, der Vet, HCW 2012, 'Missing Data and Imputation: A Practical Illustration in a Prognostic Study on Low Back Pain', *Journal of manupulative and therapeutics*, vol. 35, no. 6, pp. 464–471, doi: 10.1016/j.jmpt.2012.07.002.

Virginia, R, de Albuquerque, C & Heller, L 2018, 'The human rights to water and sanitation: challenges and implications for future priorities' in *In Equality in Water and Sanitation Services,* Routledge*,* pp. 42-61

Vittinghoff, E & McCulloch, CE 2007, 'Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression', *American Journal of Epidemiology,* vol. 165, no. 6, pp. 710–718, https://doi.org/10.1093/aje/kwk052.

WHO 2009, *World Health Organization - Global health risks: mortality and burden of disease attributable to selected major risks,* viewed Febuary 12 2018, https://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf
Wikipedia contributors 2018, *Lake Bangweulu*, viewed May 25, 2018 from https://en.wikipedia.org/w/index.php?title=Lake_Bangweulu&oldid=838396671

World Health Organization 2017, *United Nations Children's Fund (UNICEF). Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and Sdg Baselines*, viewed, Febuary 12 2018, http://41.77.4.165:6510/www.who.int/mediacentre/news/releases/2017/launch-version-report-jmp-water-sanitation-hygiene.pdf

Zimba, R, Ngulube, V, Lukama, C, Manangi, A, Osbert, N, Hoehne, A, Muleya, S, Mukosha, L, Crooks, P, Chikobo, C, Winters, B, Larsen DA 2016, 'Chiengi District, Zambia Open Defecation Free After 1 Year of Community-Led Total Sanitation', *The American Journal of Tropical Medicine and Hygiene, vol. 95*, no. 4, pp. 925-927, doi:10.4269/ajtmh.16-0210.

9 APPENDIX

Examples 1 to 5 provide illustrations on the practical interpretation of the prognostic model used the developed calculator. Using the tool, downloadable here, the values in the model are inputted to generate the following results:

$$P = \frac{1}{1+e^{-x}} \tag{1}$$

where;
$x = -4.459\ (0.245t + 0.01n + 0.112\ell - 1.599h + 1.540v)$
P = ODF status loss (adequate latrines)
t = time (months)
n = village population (people)
ℓ = number of latrines built after CLTS (latrines)
h = latrines with handwashing with soap facility (handwashing facility)
v = latrine privacy (latrine wall and door or suitable acceptable substitutes)

Example 1: Month following ODF status attainment
A village within the first month (month = 0) of ODF status and a village population of 105 people (17 households) with 17 latrines built after CLTS, 17 households with handwashing with soap facilities and 17 households with latrines providing privacy, when presented with a risk of ODF status loss, would have the risk of status loss of 7%. Maintaining all prognostic factors constant, an increase in the number of months after ODF attainment to 6 months, increases the risk of ODF status loss to 25%. Whilst an increase to 12 months, the risk of ODF status loss increases to 60%.

Scenario 1	Scenario 2	Scenario 3
$t = 0$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$	$t = 6$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$	$t = 12$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$

Example 2: Village population
A village with 5 months ODF status having a village population of 60 people (10 households) with 10 latrines built after CLTS, 10 households with handwashing with soap facilities and 10 households with latrines providing privacy, when presented with a risk of ODF status loss, would have the risk of status loss of 11%. Maintaining all prognostic factors constant, an increase in village population to 120 (20 households), increases the risk of ODF status loss to 18%. Whilst a threefold increase to 180, will increase the ODF status loss to 28%.

Scenario 1	Scenario 2	Scenario 3
$t = 5$ $n = 60$ $\ell = 10$ $h = 10$ $v = 10$	$t = 6$ $n = 120$ $\ell = 17$ $h = 17$ $v = 17$	$t = 12$ $n = 180$ $\ell = 17$ $h = 17$ $v = 17$

Example 3: Latrines built after a CLTS intervention
A village with 5 months ODF status having a village population of 105 people (17 households) with 0 latrines built after CLTS, 17 households with handwashing with soap facilities and 17 households with latrines providing privacy, when presented with a risk of ODF status loss, would have the risk of status loss of 4%. Maintaining all prognostic factors constant, an increase in latrines built after a CLTS intervention by half (8 of 17 households), increases the risk of ODF status loss to 9%. When all the 17 households in the village have all their latrines built after a CLTS intervention, the risk of ODF status loss increases to 21%.

Scenario 1	Scenario 2	Scenario 3
$t = 5$ $n = 105$ $\ell = 0$ $h = 17$ $v = 17$	$t = 5$ $n = 105$ $\ell = 8$ $h = 17$ $v = 17$	$t = 5$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$

Example 4: Handwashing with soap facilities

A village with 5 months ODF status having a village population of 105 people (17 households) with 17 latrines built after CLTS, and none of the households with handwashing with soap facilities, whilst all the 17 households are with latrines providing privacy, when presented with a risk of ODF status loss, would have the risk of status loss of 100%. Maintaining all prognostic factors constant, an increase the households with latrines having handwashing with soap facilities by half (8 of 17 households), still maintains the risk of ODF status loss at 100%. When all the 17 households in the village have all their latrines with a handwashing with soap facility, the risk of ODF status loss reduces to 21%.

Scenario 1	Scenario 2	Scenario 3
$t = 5$ $n = 105$ $\ell = 17$ $h = 0$ $v = 17$	$t = 5$ $n = 105$ $\ell = 17$ $h = 8$ $v = 17$	$t = 5$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$

Example 5: Latrines providing privacy
A village with 5 months ODF status having a village population of 105 people (17 households) with 17 latrines built after CLTS, and 17 households with handwashing with soap facilities, whilst all none of the households are with latrines providing privacy, when presented with a risk of ODF status loss, would have the risk of status loss at several multiples of a 100%. Maintaining all prognostic factors constant, an increase the households with latrines providing privacy (8 of 17 households), still maintains the risk of ODF at several multiples of a 100%. When all the 17 households in the village have all their latrines providing privacy, the risk of ODF status loss reduces to 21%.

Scenario 1	Scenario 2	Scenario 3
$t = 5$ $n = 105$ $\ell = 17$ $h = 17$ $v = 0$	$t = 5$ $n = 105$ $\ell = 17$ $h = 17$ $v = 8$	$t = 5$ $n = 105$ $\ell = 17$ $h = 17$ $v = 17$