Exploring Factors Influencing Water Accessibility in Liberia in Regression Modeling.

GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

UNC WATER INSTITUTE

Henry Li [1], Catherine McManus[2],Ryan Cronk PhD[2],

1. Department of Biostatistics Gillings School of Global Public Health, University of North Carolina, Chapel Hill
2. Department of Environmental Sciences and Engineering Gillings School of Global Public Health, University of North Carolina, Chapel Hill

Background

- **Water Access issue is critical and exacerbated by Climate Change:** 25% of the global population lacks access to safely managed drinking water services.[1]
- **Countries In Sub-Saharan Africa are affected by the most:** Liberia, in particular struggles with clean water access: 10% of its population having access to safe drinking water.[2]
- **UN's WASH Initiative and survey :** The United Nations launched the "Water, Sanitation and Hygiene (WASH) for All" initiative. Collaborating with the Liberia government which include a comprehensive survey of water points. (***N=20205***).
- Insights from the WASH water point survey can improve and expand clean water access by providing insights with governing and policy changes regarding WASH services.

Purpose

- To use modeling approach to navigate through complex systems on Water delivery system
- To evaluate the performance of hand pumps and factors that may improve their functionality/damage status.

Method

- The WASH dataset from WASH Liberia's website was downloaded and cleaned using R.
- Categorize the WASH survey questions and responses based on:
  - **Geographical information**
  - **Functionality information**
  - **Type of water point and pump data**
  - **Damage information**
  - **Water point reliability**
  - **Installer, maintainer, owner, and fee status.**
- The Entire dataset (**N=20205**) with improved water points (**N=13239**) was subset to observations having Afridev manual handpumps only (**N=11796**).
- Missing data was imputed using SAS (**Statistical Analysis System**)'s *Proc surveyimpute procedure (~10%).*
  - Via **Hot Deck Bayesian Bootstrapping method**.
- Descriptive statistics and select χ^2 statistics values are conducted.
- A multivariate logistic regression model and an ordinal logistic regression model was built with the intention to evaluate
  - Functionality.
  - Reliability for functioning waterpoints.
- Model Assessment and validation and performed using SAS.
- Variables are selected a-priori and through backward method.

Results

Functionality – of all Afridev branded hand pumps**:**

- **80.61%** are considered **functional (N=9509)**
- **19.39%** are considered **non-functional (N=2287)**

Reliability - out of all functional Afridev branded pumps**:**

- **90.6%** are considered **reliable** (almost always) **(N=8262)**
- **5.5%** are considered **reasonable** (<7 days unavailable a week) **(N=505)**
- **3.9%** are considered **insufficient** (>7 days unavailable a month **(N=353)**

Survey Question		Functional (N=9529)	Non-Functional (N=2287)	p-value**
Who maintains the water point?				
	-WASH comittee	2881 (78.18%)	804 (21.82%)	**<0.0001**
	- Other	6628 (81.72%)	1483 (18.28%)	
Is there a trained mechanic available for this water point? Was the mechanic provided with toolkits ?				
	- No Mechanic	5075 (79.83%)	1282 (20.17%)	**<0.0001**
	- Mechanic witout toolkits	766 (75.17%)	253 (24.83%)	
	- Mechanic with toolkits	3668 (82.99%)	752 (17.01%)	
Was there damages on the water point?				
	- Yes	2508 (56.65%)	1919 (43.35%)	**<0.0001**
	- No	7000 (95.01%)	368 (4.99%)	
Last time the water point broke down, how long did it take to repair?				
	- Never Broken	3877 (94.82%)	212 (5.18%)	**0.045**
	- More than a year	950 (48.79%)	1038 (52.21%)	
	- Over a month	2276 (75.02%)	758 (24.98%)	
	- Over a week	1168 (87.43%)	168 (12.57%)	
	- Less than a week	1238 (91.77%)	111 (8.23%)	

** Mantel-Haenszel Chi-Square

Table 1. Descriptive statistics of response to survey functionality question per group.

Based on the distribution and cross tabulation, a logistic model using functionality as a binary response variable was conducted.

- **Response variable**: Functionality (Yes/No)
- **Predictor variable**:
  - WASH management
  - Mechanic and toolkit availability
  - Damage status
  - Repair time
- AUC: 0.8043

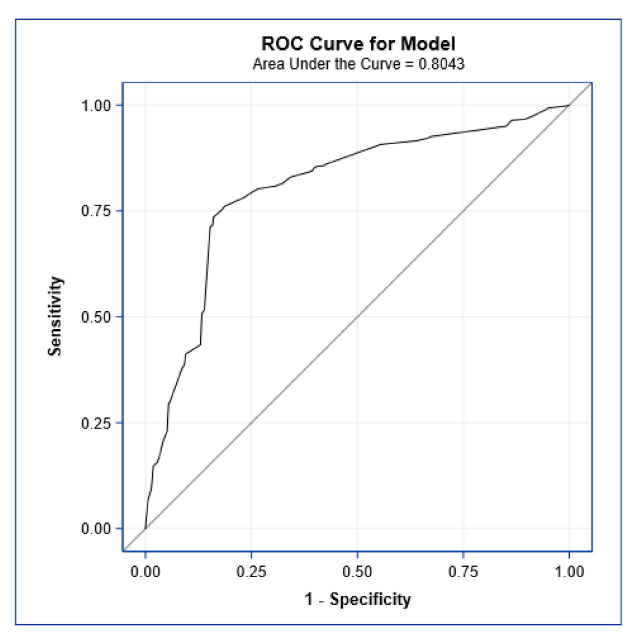

Figure 1. AUC for logistic model

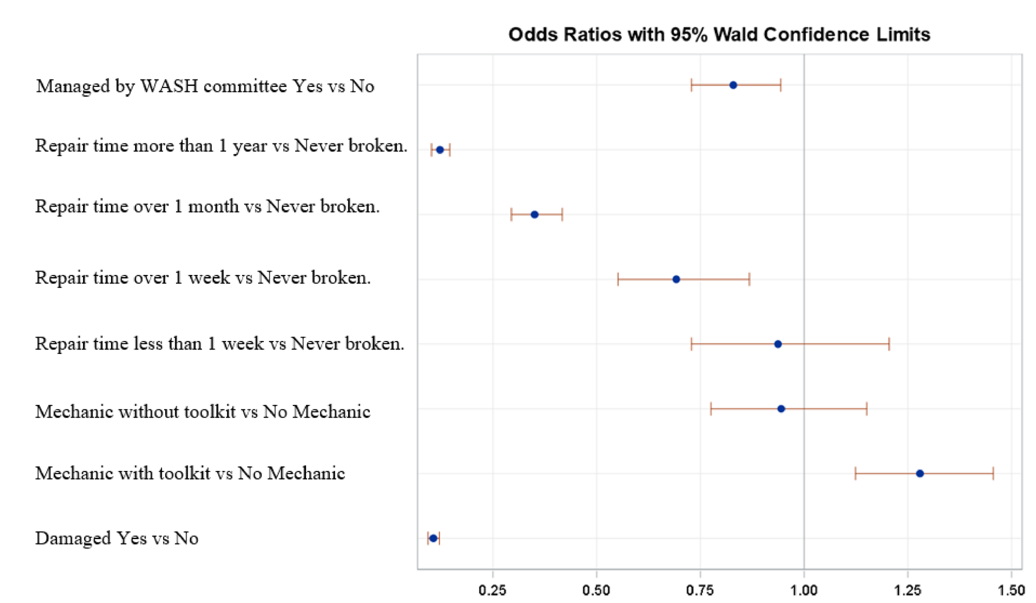

Figure 2. Odds ratio for each variable in model - Functionality.

Explanatory variable	OR	CI	p-value **
Damage status: Yes vs No	0.107	0.104-0.094	**<0.0001**
Repair time over a month vs never broken	0.351	0.351-0.295	**<0.0001**
Repair time over a week vs never broken	0.692	0.692-0.551	**0.0015**
Repair time more than a year vs never broken	0.122	0.102-0.146	**<0.0001**
Mechanic with toolkit vs No Mechanic	1.279	1.279-1.124	**0.0005**
Repair time less than a week vs never broken	0.937	0.391-0.729	0.6110
Mechanic without toolkit vs no mechanic	0.945	0.945-0.776	0.5715
Managed by WASH comittee Yes vs No	0.829	0.729-0.943	0.0045

Table 2. Odds ratio of functional status for each variable. ** Wald-type OR χ^2 p-value

Results

Survey Question		Reliable (N=8262)	Reasonable (N=505)	Insufficient (N=353)	p-value **
Who maintains the water point?					
	- WASH comittee	2588 (92.89%)	74 (2.66%)	124 (4.45%)	**<0.0001**
	- Other	5680 (89.59%)	279 (4.4%)	381 (6.01%)	
Is there a trained mechanic available for this water point? If so, was the mechanic provided with toolkits ?					
	- No Mechanic	4303 (89.07%)	237 (4.91%)	291 (6.02%)	**<0.0001**
	- Mechanic without toolkits	651 (87.97%)	36 (4.86%)	53 (7.16%)	
	- Mechanic with toolkits	3314 (2.25)	80 (2.25%)	161 (4.53%)	
Was there damages on the water point?					
	- Yes	1871 (80.68%)	187 (8.06%)	261 (11.3%)	**<0.0001**
	- No	6396 (93.98%)	166 (2.44%)	244 (3.59%)	
Is the water paid for at this point?					
	- No, It's Free	7236 (91.33%)	289 (3.65%)	398 (5.02%)	**<0.0001**
	- Only after a system breakdown	583 (85.36%)	43 (6.3%)	57 (8.35%)	
	- Yes, a flat fee	276 (83.13%)	14(4.22%)	42 (12.65%)	
	- Yes, by volume	173 (92.02%)	7 (3.72%)	8 (4.26%)	

** Mantel-Haenszel Chi-Square

Table 3. Descriptive statistics of response to reliability survey question per group.

Explanatory variable	OR	CI	p-value **
Damage status: Yes vs No	0.11	0.10-0.09	**<0.0001**
Repair time over a month vs never broken	0.35	0.35-0.30	**<0.0001**
Repair time over a week vs never broken	0.69	0.69-0.55	**0.0015**
Repair time more than a year vs never broken	0.12	0.10-0.15	**<0.0001**
Mechanic with toolkit vs No Mechanic	1.28	1.28-1.12	**0.0005**
Repair time less than a week vs never broken	0.94	0.39-0.73	0.6110
Mechanic without toolkit vs no mechanic	0.95	0.95-0.78	0.5715
Managed by WASH comittee Yes vs No	0.83	0.73-0.94	0.0045

** Wald-type OR χ^2 p-value

Table 4. Odds ratio of functional status for each variable in model 2.

Conclusions

- Undamaged water systems and systems that were repaired within a week were strongly associated with higher rates of functionality (table 2).
- Water points managed by WASH committee is more reliable/marginally functional than other parties.
- Providing mechanic with toolkits is more important than their presence.
- Functionality model can distinguish between positive and negative cases with an accuracy of ~ 80%.
- Reliability model has prediction accuracy ~74%

Limitations

- Bootsatrping imputation of missing value (~10% of all) has assumption of missing (completely) at random.
- Reliability model may only inference to functioning waterpoints due to survey.
- Small sample size on several variables.
- Non-probability sample survey requires adjustment from prior knowledge

Future Perspective

- Using a Bayesian network-based inference to further explore factors influencing water point functionality and reliability
- Continue to improve the performance of the logistic ordinal model

References