

## PRE-ANALYSIS PLAN

Effect of a Community-Driven Water, Sanitation, and Hygiene  
Program on Covid-19 knowledge, behavior, and health:  
A Randomized Control Trial in the Democratic Republic of Congo

July 2021

Aidan Coville	Kevin Croke	Eric Mvukiyehe	John Quattrochi
Research Department	School of Public Health	Dept. of Political Science	Dept. of Public Health
The World Bank	Harvard University	Duke University	Simmons University

## Introduction

In 2008, the Government of the DRC launched a national program known as National Healthy Villages and Schools (in French "*Villages et Ecoles Assainis*", VEA), a water and sanitation program financed by DFID and implemented by UNICEF and the Government of DRC's Ministry of Public Health and Ministry of Primary, Secondary, and Professional Education. The VEA's main objectives are to support communities' improved access to water, hygiene, and sanitation services and practices through the construction of water and sanitation infrastructure in villages and schools, local community involvement, and a village certification mechanism.

This study uses a cluster-level randomized controlled trial to evaluate the causal impact of the VEA program across 332 rural villages, with a focus on COVID-related outcomes. The VEA program was implemented in these villages starting in 2019. A main evaluation involves experimentally varying the provision of the VEA program to a set of villages to examine the causal impacts on water access, infrastructure development, availability of sanitation services, knowledge of hygiene practices, and child health. An first paper from this project assesses the short run effects of the program at approximately 5 months post-intervention

(<https://www.socialscisearch.org/trials/4648/history/64498>). However, as this evaluation was in process, the COVID-19 pandemic occurred. The research team therefore added a research component to study the impact of VEA on COVID related outcomes, to be measured through mobile phone interviews. This pre-analysis plan therefore pertains to 12, 18, and 24 month follow-up surveys, implemented through mobile phone surveying modalities, to measure VEA program effects on mental and physical health, Covid-19 knowledge and prevention, and governance.

The first mobile phone survey was implemented on an emergency basis in the first phase of COVID-19 in DRC, so a pre-analysis plan was not prepared for round 1 data. Because of this, there are no R1 outcomes included in the study's primary outcomes. The pre-analysis plan was created and filed before round 2 and 3 data had been collected and analyzed. Minor revisions (detailed below) were made after the authors had analyzed data from the second round of the phone survey, before they had seen any data from the third round of the phone survey. Data from the third round was collected in June – July 2021, during a period when reported COVID cases were rising to their highest levels to date in DRC.

## Research questions

1. What is the effect of community-led WASH on Covid-19 illness, other illnesses, child health, and psychological well-being?

2. What is the effect of community-led WASH on Covid-19 knowledge, Covid-19 prevention, governance, livelihoods, food security, and access to health care?

### Timeline

June 2019: Median month at which program implementation was completed

Nov-Dec 2019: Household survey

May-July 2020: Round 1 phone survey

Nov 2020-Jan 2021: Round 2 phone survey

June-July 2021: Round 3 phone survey

### Outcomes

We list the primary and secondary outcomes in Table 1. For outcomes measured in multiple rounds, we will calculate the mean across rounds. We have a total of 5 primary and 22 secondary outcomes.

**Table 1. Primary and secondary outcomes**

Outcome Type*	Outcome Name	Definition	Survey round(s)
Primary	Covid symptoms	Number of household members in past week with fever, dry cough, difficulty breathing/shortness of breath, or fatigue	2 and 3
Primary	Non-Covid illnesses	Number of sick household members in last seven days [exclude those with Covid symptoms]	2 and 3
Primary	Child health index	Summary index of proportion of children under five with fever/cough/diarrhea in last two weeks [this will be three separate numbers for each HH with children under 5]	2 and 3
Primary	Psychological well-being	Summary index of scores to following questions (1=All of the time...6=None of the time):  Have you been a very nervous person over the past four weeks ?	2

		<p>Have you felt so down in the dumps that nothing could cheer you up over the past four weeks ?</p> <p>Have you felt calm and peaceful over the past four weeks ? (negatively coded)</p> <p>Have you felt downhearted and blue over the past four weeks ?</p> <p>Have you been a happy person over the past four weeks ? (negatively coded)</p>	
Primary	Vaccine acceptance	If a vaccine for COVID-19 becomes available to you, would you take it?	3
Secondary	Village Covid prevention	Summary index of the following binary variables as reported by households: banned gatherings; community meetings; HH visits; curfew; closed shops; handwashing materials/water; isolation of sick; medical care; wear mask; social distancing	1 and 2
Secondary	Household Covid prevention	Summary index of the following binary variables from household survey [Have you practiced X in the last two weeks?]: wash hands frequently; avoid physical contact; avoid social gatherings; stay at home; wear mask; keep 1+ meter distance	1 and 2
Secondary	Perceptions of Covid prevention	Summary index of the following binary variables from household survey [Have most people in your village/community practiced X in the last two weeks?]: wash hands frequently; avoid physical contact; avoid social gatherings; stay at home; wear mask; keep 1+ meter distance	2
Secondary	Respondent Covid knowledge	Summary index of the following binary variables from household survey [Is X effective to prevent Covid?]: wash hands frequently; avoid physical contact; avoid social gatherings; stay at home; wear mask; keep 1+ meter distance; inject/inhale disinfectant (negative)	1
Secondary	Covid Governance	<p>Summary index of Likert variables:</p> <p>A23. Overall, how well do you think the government is doing in its response to the COVID epidemic? [1=very badly...5=very good]</p>	1

		A24. How much do you trust the government to respond well to the COVID epidemic? [1=strongly distrust...5=strongly trust]	
Secondary	Livelihoods	<p>Summary index of responses to:</p> <p>A27.a In the past week, how many hours have you spent working on an income generating activity?</p> <p>A27.c In the past week, how many hours has your spouse/partner spent working on an income generating activity for your household? [if applicable]</p> <p>A27.d Has anyone in your household (including you) lost their job (or main source of income) since the coronavirus pandemic began? [negatively coded]</p> <p>A27.e Have you or anyone in your household gained a new job (or main source of income) since the coronavirus pandemic began?</p> <p>A27.f Think of the income you earned each month before the coronavirus pandemic began. How does that compare to the income you are earning now?</p>	2 and 3
Secondary	Food security	In the last seven days, how many days (out of 7) did you lack enough food to eat?	1, 2, and 3
Secondary	Access to health care	Have you or any other person your household delayed, skipped or been unable to complete health care visits since the Coronavirus pandemic began?	2
Secondary	Hospital visits	Number of household members who visited hospital in last seven days	1, 2, and 3
Secondary	Household illnesses	Number of sick household members in last seven days [includes those with Covid symptoms, because no data on symptoms was collected]	1
Secondary	Approval of authorities (10 outcomes)	<p><i>Each authority listed below is a separate outcome</i></p> <p>Do you approve or disapprove of the way that the following people have performed their jobs over the past twelve months, or haven't you heard enough about them to say?</p> <ol style="list-style-type: none"> <li>1. President</li> <li>2. National Assembly</li> <li>3. Ministry of Health</li> </ol>	3

		<ol style="list-style-type: none"> <li>4. Provincial Government</li> <li>5. International NGOs</li> <li>6. Local NGOs</li> <li>7. Traditional leaders</li> <li>8. Health zone officials</li> <li>9. Health area officials</li> <li>10. Village chief</li> </ol>	
Secondary	Leader vaccine acceptance	<i>In village leader survey:</i> If a vaccine for COVID-19 becomes available to you, would you take it?	3
Secondary	Leader vaccine advice	<i>In village leader survey:</i> Will you advise the people in your village to take the Covid vaccine?	3

## Estimation

We will separately estimate the main impacts of VEA on the outcomes listed above, using the following basic specification:

$$y_{ihvc} = \alpha + \beta_1 T_c + \gamma X_{ihvc} + \delta Z_{ihvc} + \varepsilon_i \quad (1)$$

where  $y_{ihvc}$  is the outcome of interest for respondent  $i$  in household  $h$  in village  $v$  in cluster  $c$  at the follow-up survey, defined above.  $T_c$  is the treatment indicator that takes value 1 for clusters that were randomly assigned to participate in VEA (“treatment clusters”) and 0 for otherwise (“control clusters”).  $Z_{ihvc}$  represents a set of strata-specific dummies where strata are based on province and number of villages in the cluster, which will equal 1 if the household falls in that stratum, and 0 otherwise.  $X_{ihvc}$  is a vector of baseline covariates included in the analysis. Specifically we include gender and age (month) dummies for all <5 child health outcomes, but do not include covariates for other outcomes.  $\delta$  and  $\gamma$  are vectors of associated strata and covariate coefficients respectively.  $\varepsilon_i$  is an idiosyncratic error term. Our main parameter of interest is  $\beta_1$ , the intention-to-treat effect (ITT). Standard errors will be clustered at the randomization (village cluster) level.

To reduce the number of statistical tests and reduce the probability of false positives (Type I errors), when an outcome space is comprised of multiple variables (such as the COVID prevention behavior outcomes) we combine measures into an index following (Kling, Liebman, and Katz 2007). We will use the “greedy” version of the index that does not impute for missing values.

## Attrition

If we detect statistically significant differences in survey attrition by treatment status, we will use inverse probability weighting to account for this differential attrition by treatment status. We will measure attrition at the household level. In the following, A=treatment, L=control variables, and C=censoring, i.e. attrition

1.  $f(A|L)$ : Predicted values from logistic regression of treatment on randomization clusters and baseline characteristics (variables in balance test)
2.  $f(A)$ : Predicted values from logistic regression of treatment
3.  $f(C|L)$ : Predicted values from logistic regression of attrition on treatment, randomization clusters, and baseline characteristics (variables in balance test)
4.  $f(C)$ : Predicted values from logistic regression of attrition on treatment
5. Treatment weight =  $f(A)/f(A|L)$
6. Censoring weight =  $f(C)/f(C|L)$
7. Final weight = treatment weight \* censoring weight
8. Fit model with final weight as probability weight

## Balance test

Following our analysis of the previous round of data collection, we will include the following variables in a balance test:

- Type of roof (improved roof= 1 if roof is finished roofing (i.e. Metal, Wood, Calamine/Cement fiber Ceramic tiles, Cement, or Roofing shingles))
- Wall (improved walls = 1 if walls are “finished walls”)
- Floor (improved floor= 1 if floor is “finished floor”)
- Household size
- Respondent religion (Catholic, Protestant, or Other)
- Respondent age
- Respondent education (less than primary; completed primary; completed secondary)
- Marital status (Married/cohabitating vs not)

## Subgroup analyses

As an exploratory analysis, we will test for differences in effect across provinces.

## Changes to this pre-analysis plan

We made the following changes before seeing the data from Round 3, but after seeing data from Rounds 1 and 2:

- Changed mental health to psychological well-being
- Changed “For outcomes measured in both round 1 and 2, we consider each round to be a separate outcome” to “For outcomes measured in multiple rounds, we will calculate the mean across rounds.”

We added the following primary outcome:

1. Vaccine acceptance

We added the following 12 secondary outcomes:

1. Approval of President
2. Approval of National Assembly
3. Approval of Ministry of Health
4. Approval of Provincial Government
5. Approval of International NGOs
6. Approval of Local NGOs
7. Approval of Traditional leaders
8. Approval of Health zone officials
9. Approval of Health area officials
10. Approval of Village chief
11. Village leader Covid vaccine willingness
12. Village leader advice on Covid vaccine