Statistical Analysis Plan

Version 1.1

Integration of Serial Self-Testing into Public Health Contact Tracing Programs: A Pragmatic Trial to Assess the Operational Feasibility and Impact of COVID-19 Self-Testing among Exposed Individuals in Brazil

Protocol Number: 1889025-1

Document Version History

Version Date	Version	Author	Change Description	Reason/Comment
01 October 2023	1.1	Rebecca Green	Updates to GPS analysis.	Refinement of exploratory GPS objectives given large amount of missing data
03 March 2023	1.0	Rebecca Green	Initial release.	Not applicable.

TABLE OF CONTENTS

1. STUDY OBJECTIVES

1.1. PRIMARY OBJECTIVE

To evaluate the effectiveness of contact tracing supported by serial self-testing (testing daily for up to 10 days) among exposed individuals compared to contact tracing using only professional testing with Ag-RDT performed at one visit.

1.2. ANALYTICAL SECONDARY OBJECTIVES

- To evaluate the operational feasibility of self-testing within the public health contact tracing system.
- To evaluate the concordance of supervised self-tests as compared to an Ag-RDT conducted by a trained health worker.

2. BACKGROUND/INTRODUCTION

See protocol for detailed description of study background and design. In brief, close contacts of index cases testing positive for COVID-19 will be enrolled into one of two arms. The control arm will receive standard contact tracing with a professional test at one time point post-exposure. The intervention arm will additionally receive 10 self-tests to be completed over the follow-up period. Randomization will occur at the index case level. Participants will be followed up for 10 days.

2.1. SAMPLE SIZE

This study is powered to demonstrate a 7.5% difference in positive cases identified between Arm 1 and Arm 2. Prior work at this site with COVID testing of close contacts yielded a 30% PCR positivity rate among close contacts during a period of low to moderate transmission. To account for increases in vaccination coverage, high rates of prior infection, and the likelihood of low transmission following the Omicron wave, we estimate this population will have a 20% test positivity rate. Based on the established performance characteristics of Ag-RDT tests, we estimate serial self-testing will identify up to 75% of those cases ($p1 = 0.15 \therefore p0 = 0.075$).

Utilizing Equation 1 listed below, where $z\alpha/2 = 1.96$ and $z\beta = 0.842$, we calculate a total of 550 participants needed to complete the study (275 per Arm). To account for attrition in longitudinal data and the exclusion of unevaluable cases, we will increase the sample size estimate by 10% to enroll a total of 604 close contacts (302 per Arm). To achieve this, we anticipate needing to enroll approximately 150 index cases, with each index case yielding an average of 4 close contacts (75 per Arm), though enrollment of index cases will continue until the desired number of close contacts is enrolled.

Equation 1: Sample Size Calculation for a Difference in Proportions

$$N = \frac{2\,(z\alpha/2 + z\beta)2 * (p0(1 - p0) + p1(1 - p1))}{\Delta}$$

Key Assumptions:

a. A reasonable positivity delta between arms is 7.5%
b. Close contact test positivity will be 20%
c. Participant attrition will be 10%
d. Each index case will yield four close contacts on average

3. POPULATIONS OF ANALYSIS

Index Cases: All participants with enroll_type = 0 and submitting at least 3 surveys during the follow-up period.

Modified Intention-to-Test (mITT): enroll_type = 1 AND NOT known COVID positive
IF ppt_arm=1, !is.na(prof_rdt_result)
IF ppt_arm=2, !is.na(st_result_cc2) for 3/10 follow-up surveys

Per Protocol: enroll_type = 1 AND
IF ppt_arm=1, !is.na(prof_rdt_result)
IF ppt_arm=2, !is.na(st_result_cc2) for 3/10 follow-up surveys

4. OUTCOME VARIABLES

Variable Class	Variable Subclass	Variable	Variable Coding	Variable Definition/Purpose
Primary Outcome	Dependent Variable	prof_rdt_result st_result_cc2	1 = Positive 0 = Negative 2 = Invalid 99 = I don't know	Professional RDT result at enrollment (Arm 1) and ST result over follow-up period (Arm 2) – to be counted as positive if any positive result is indicated over follow-up period
Primary Outcome	Independent Variable	enroll_type	0 = Index Case 1 = Close Contact ppt_arm: 1 = Control 2 = Intervention	To distinguish between index cases and close contacts as well as intervention assignment
Secondary Outcome		prof_rdt_result st_rdt_result_staff	1 = Positive 0 = Negative 99 = Invalid	Professional and ST RDT results at enrollment (Arm 2 only, for concordance)
Secondary Outcome		gps_success_yn	1 = Yes 0 = No	GPS data sharing successful
Administrative		enroll_date	Follow-Up Window = enroll_date + 10	Calculate participant follow-up window
Covariates	Demographic Information	dob and enroll_date	Age = enroll_date - dob	Calculate participant age
		sex	0 = Male 1 = Female	Participant sex
		race	1 = African 2 = Asian 3 = Caucasian 4 = Indigenous 5 = Other	Participant race
		education	0 = Illiterate -1 = Incomplete Primary Education 1 = Completed Primary Education	Participant education

			-2 = Incomplete Secondary Education 2 = Completed Secondary Education -3 Incomplete Tertiary Education 3 = Completed Tertiary Education 4 = Other	
		marital_status	1 = Single 2 = Married 3 = Divorced 4 = Widowed 5 = Partnered (have a significant other) 6 = Other 999 = Prefer not to say	Participant marital status
	Participant Risk Factors	hh_size	Integer	Number of people living in participant's household (excluding participant)
		live_with_old_yn live_with_young_yn	1 = Yes 0 = No	Participant household risk factor
		work_outside_home	0 = None of the time 1 = Some of the time 2 = Most of the time 3 = All of the time	How often participant works outside of their home
		chr_resp_dis copd chr_kidn_dis cancer immcomp chr_hrt_dis diabetes pregnant postpartum obese severe_obese smoker hyperten sickle_cell	1 = Yes 0 = No 999 = Unknown	Participant's pre-existing conditions
		covax_yn covax_doses	1 = Yes 0 = No integer	COVID Vaccination Status
		sxs_enroll	(Check all that apply) (1) Fever or Chills (2) Cough (3) Shortness of breath or difficulty breathing (4) Fatigue	Symptoms present at time of enrollment

			(5) Muscle or body aches (6) Headache (7) New loss of taste or smell (8) Sore throat (9) Congestion or runny nose (10) Nausea or vomiting (11) Diarrhea (12) Other (13) None of the above	

5. STATISTICAL METHODOLOGY

5.1. GENERAL METHODOLOGY

Participant characteristics will be compared between arms and sites in a similar fashion to the tables below; italicized variables denote continuous measures and bolded variables denote categorical measures. Statistically significant differences between groups will be noted if $p<0.05$ and covariates will be included in the regression model if $p<0.05$ between Arms.

Table 1: Close Contact Population Characteristics Overall and by Site

	Overall	**Porto Velho**	**Curitiba**
N			
Age			
Sex			
Race			
Education			
Marital Status			
Household Size			
Live with young			
Live with old			
Frequency of work outside home			
Household Risk Status Composite (3-point scale with 1 point per lives with young, lives with old, and works outside the home most/all of time)			
1+ Pre-existing Condition			
Received COVID Vaccination			

Table 1a: Index Case Population Characteristics Overall and by Site

	Overall	**Porto Velho**	**Curitiba**
N			
Age			
Sex			
Race			
Education			
Marital Status			
Received COVID Vaccination			

Table 2: Close Contact Population Characteristics between Arms

	Overall	**Control**	**Intervention**
N			
Age			
Sex			
Race			
Education			
Marital Status			
Household Size			
Live with young			
Live with old			
Frequency of work outside home			
Household Risk Status Composite (3-point scale with 1 point per lives with young, lives with old, and works outside the home most/all of time)			
1+ Pre-existing Condition			
Received COVID Vaccination			

Table 2a: Index Case Population Characteristics between Arms

	Overall	**Control**	**Intervention**
N			
Age			
Sex			
Race			
Education			
Marital Status			
Received COVID Vaccination			

HANDLING OF MISSING DATA

Only complete cases will be included; missing data will not be imputed. For the primary analysis, complete cases are defined for Arm 1 as having a professional RDT result available at enrolment and for Arm 2 as having submitted at least three self-test results over the ten-day follow-up.

SUBGROUP ANALYSIS

No subgroup analysis will be performed for the primary analysis. Secondary analyses will contain analyses by subgroup. Specifically for positive cases, endpoints regarding adherence to public health guidelines such as intra-household isolation and masking behaviour will be stratified by sex, household risk status, and presence of 1+ pre-existing condition.

5.2. INTERIM ANALYSIS

An informal interim analysis will be conducted with the first 250 close contacts who have completed the follow-up period to check the Key Assumptions for the sample size calculation. No statistical comparisons will be made. The analytical information assessed will include:

- Overall close contact positivity rate to date as defined by positive professional Ag-RDT result at enrollment.
- Test positivity rate of each arm as defined by:
    - Arm 1: positive professional Ag-RDT result on Day 0 [enrollment]
    - Arm 2: positive professional Ag-RDT result on Day 0 [enrollment] OR positive self-test result during follow-up [D1-D10]

- Rate of attrition (i.e., incomplete cases) as defined by number of close contacts who have completed < 3 surveys during the follow-up period.
- Average number of close contacts yielded per index case.

Upon generation of the above information, the following will be performed to inform subsequent study planning activities:

- The sample size calculation will be performed again using the updated parameters as described in section 2.1.
    - Outcome: to inform investigators of whether the initial sample size calculation is still valid; this will guide decisions around protocol amendments to increase the number of participants and/or study timeline considerations
- The estimated number of required index cases will be calculated based on the average yield of close contacts.
    - Outcome: to inform investigators of how many more index cases may be needed and guide a decision on whether that will be feasible given the current/projected epidemiological situation

Additional supporting information will also be generated to assist with study planning efforts, including overall recruitment by week to date, close contact/index case enrollment by week to date, and positivity rate by week to date.

5.3. PRIMARY DATA ANALYSES

All analyses will be conducted in R. A binomial logistic regression will be conducted to evaluate the difference in close contacts per index case testing positive for SARS-CoV-2 infection between Arms. The regression equation will be represented as follows, where Y is the outcome (tested positive), β_0 is the intercept, X is the independent variable (arm assignment), and Z_i is each of i imbalanced covariates.

Equation 2: Logistic Regression Equation

$$Y = \beta_0 + \beta_1 X + \beta_2 Z_i$$

In R, this will be represented as outcome ~ arm assignment + $covariate_i$ using the glm function with family = binomial.

Clustering will not be performed either by site or by index case. As imbalanced characteristics between arms will already be accounted for, it is redundant to account for site clustering. Moreover, since this is a randomized trial, any index case-induced cluster effects that may be introduced are expected to be evenly distributed between arms.

If present in the model ($p<0.05$ between Arms), continuous covariates will be centred around their mean. Model fit will be verified using Person's χ^2 test of residuals and accepted with $p>0.05$.

Raw results will be exponentiated to yield odds ratios and presented based on the following table:

Table 3: Logistic Regression Analysis of the Intervention Effect

	Odds Ratio (OR) [95%CI]
Serial Self-Testing Intervention	
$Covariate_i$	
$Covariate_i$	

5.4. SECONDARY DATA ANALYSES

OPERATIONAL FEASIBILITY

Operational Performance

To calculate operational performance, device failure percentages will be calculated as the total number of invalid tests divided by the total number of tests performed and then multiplied by 100 to yield a percentage. This calculation will be performed for the professionally administered Ag-RDT results, the supervised self-test Ag-RDT results, and the unsupervised self-test Ag-RDT results.

Data Reporting System Feasibility

Proportion of participants opting for WhatsApp/survey reporting
Message send failure rate
Number of follow-up messages sent
Number of surveys completed during follow-up

CONCORDANCE

Test concordance between the supervised self-test Ag-RDT and professionally administered Ag-RDT will be evaluated. These tests are administered at the same time (alternating which is performed first based on subject ID to minimize sample depletion-induced result bias) for subjects enrolled in the intervention arm.

The data will be subset to all participants with enroll_type = 1 and ppt_arm = 2. A binary variable will be created to indicate when prof_rdt_result = st_rdt_result_staff. The percent agreement will be calculated by totaling the number of agreements divided by the total number of paired tests and multiplied by 100.

GPS DATA ANALYSIS

This is an exploratory outcome and analyses will be dependent on the quantity and quality of GPS data provided across study participants. GPS analysis will aim to assess the accuracy of self-reported survey data compared to submitted location data.