STATISTICAL ANALYSIS PLAN

CONFIDENTIAL

Protocol Title: Transmission of Covid-19 in Crowded Environments

Short Title: The TRACE Study.

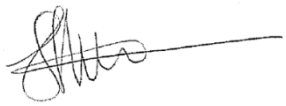
Prepared by: Sabine Hermans, Francesca Little, Robyn Curran

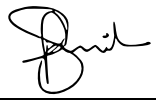
Date: 18 June 2021

SIGNATURE PAGE

STATISTICAL ANALYSIS PLAN

Version 1.0
Date 18 June 2021

Prepared by:	**Signature:**	**Date:**
Sabine Hermans Epidemiologist and leader WP5 Data Analysis Amsterdam Institute for Global Health and Development		18 June 2021
Francesca Little Professor Department of Statistical Sciences, University of Cape Town		18 June 2021
Robyn Curran Senior Researcher Desmond Tutu Health Foundation		18 June 2021

Approved by:	**Signature:**	**Date:**
Linda-Gail Bekker Principal Investigator The Desmond Tutu HIV Foundation		18 June 2021
Philip Smith Co-Principal Investigator The Desmond Tutu HIV Foundation		18 June 2021

DOCUMENT HISTORY

Version:	**Date:**	**Reason for change:**
Version 1.0	18 June 2021	Initial release

TABLE OF CONTENTS

1. INTRODUCTION

This Statistical Analysis Plan (SAP) describes the statistical methods to be used during the analysis and reporting of data collected for the TRACE study. This SAP should be read in conjunction with the Study Protocol (dated 15 December 2020), Case Report Forms (CRFs, Final Versions, dated 4 November 2020) and Data Management Plan (DMP, dated 18 June 2020).

2. OBJECTIVES OF THE STUDY

2.1. Primary objectives

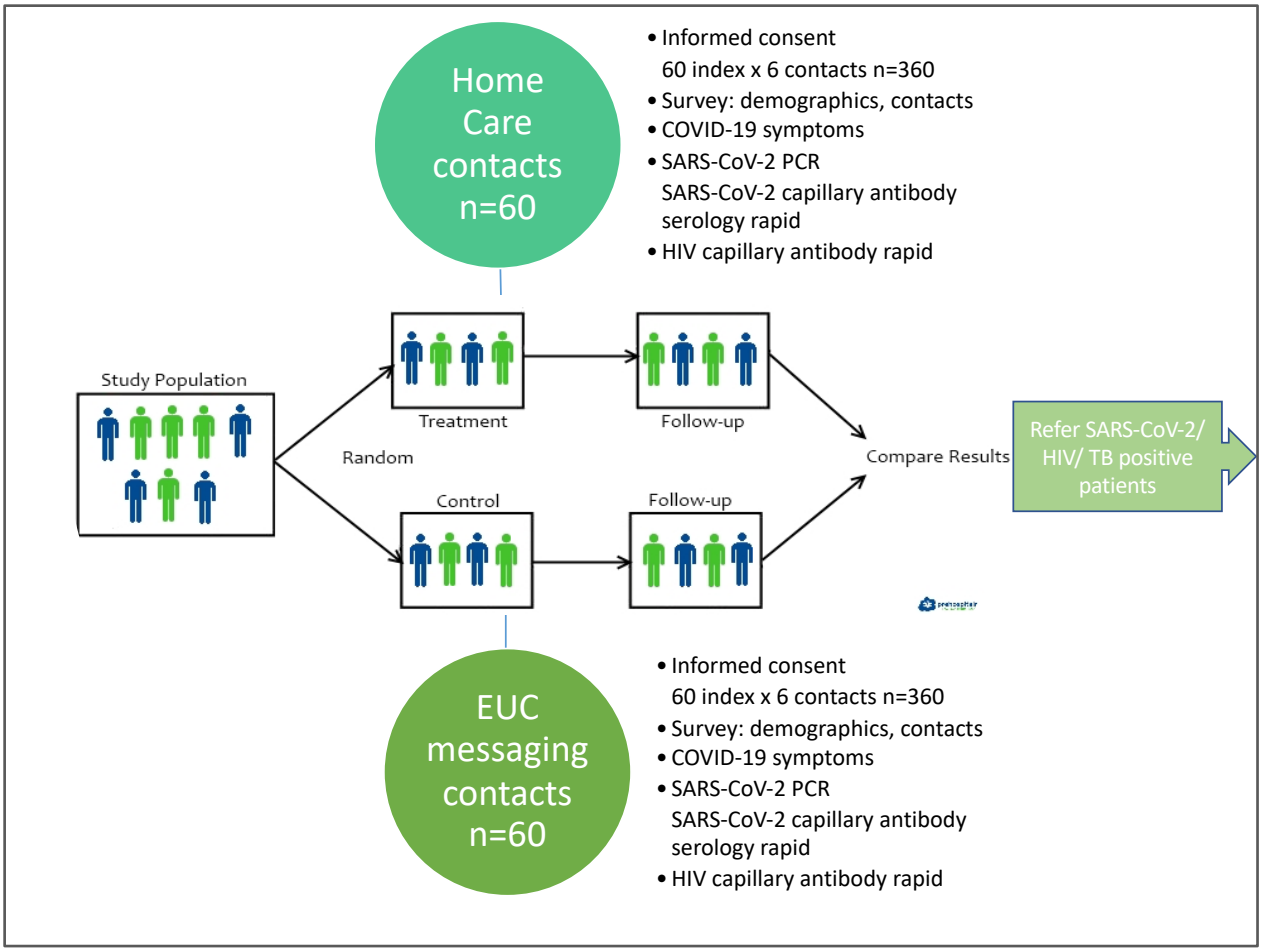
1. To measure frequency and timing of transmission of SARS-CoV-2 to household contacts.
2. To investigate the effect of an intensive infection mitigation intervention (STOPCOV) administered by lay health care workers on the household transmission of SARS-CoV-2.

2.2. Secondary objectives

1. To investigate the effect of the STOPCOV intervention on clinical outcomes and psychosocial functioning
2. To investigate the difference in transmission of SARS-CoV-2 to HHC stratified by COVID-19 wave (first or second wave)
3. To describe the household exposure to SARS-CoV-2 at baseline using the baseline serology result, and development of antibodies using the serology result during follow-up.
4. To assess acceptability, feasibility and contextual factors impacting intervention delivery (process evaluation), specifically:
   a. To understand the acceptability and feasibility of the STOPCov intervention.
   b. To understand how context shapes implementation of the STOPCov intervention at the level of households and individuals.
   c. To provide explanations for observed trends in the trial findings.
   d. To generate theoretical generalisations of how to optimise wider implementation of the STOPCov intervention to other households to reduce transmission of COVID-19.

3. DESIGN OF THE STUDY

Overview
A cluster randomised controlled study with longitudinal follow up of SARS-CoV-2 infection in 120 households which have newly diagnosed positive cases. The index cases and their HHC will be invited and enrolled after informed consent. Baseline surveys and SARs-CoV-2 screening (PCR and IgM/IgG serology) and then weekly SARs-CoV-2 screening (PCR and serology) of HHC will occur at 0, 1, 2, 3 and 4 weeks. The households will be randomized to an intensified COVID19 infection mitigation intervention (n=60) administered by CHWs vs the standard messaging (n=60). Please see Figure 1 for a schematic overview of the study design.



EUC = enhanced usual care
Figure 1: Flow chart and schematic overview of the study design.

Sample Size Determination

We calculated the different effect sizes (Table 1). Based on exponential test for comparing hazards, assuming 120*6=720 subjects, different intraclass correlation coefficients (ICCs) and comparing R0=2.5 (with resulting incidence of 0.208 among 720 subjects) to a reduced R0 of 0.9 (64% reduction, with resulting incidence of 0.075 among 720 subjects) or 1.25 (50% reduction) or 1.5 (40% reduction).

Table 1: ICC calculations

ICC	R0 2.5 to 0.9	R0 2.5 to 1.25	R0 2.5 to 1.5
	64% reduction	50% reduction	40% reduction
0.05	1.00000	1.00000	0.999944
0.25	1.00000	0.999891	0.991312
0.50	0.999993	0.995042	0.935880
0.75	0.999729	0.972942	0.846222
0.90	0.999033	0.951485	0.792745

Randomization and stratification

After completing consent, households will be randomised in a 1:1 ratio in blocks of 10 (60 households per group; 360 patient and household contacts per group).

Overview of data and data sources for analysis
All data are collected with two individual identifiers: study ID and PID (Western Cape Government health number). The exception is data source #4, which only uses PID.

1. Redcap database
    a. Baseline survey of index patient and household characteristics (demographics, COVID-19 symptoms, household tracing information, home/dwelling characteristics that impact transmission (size, rooms, ventilation, access to safe water, laundry facilities))
    b. Uptake of infection control measures at all follow-up contacts (assessment of use of isolation measures and cleaning of household surfaces up to 10 days post case of last diagnosis in household, at 1, 2, 3, 4 weeks).
    c. Assessment data of psychosocial functioning (General Health Questionnaire 9 for mood; UCLA Loneliness Scale and a version of the Brief AIDS-Related Stigma scale adapted for COVID-19, at baseline and at week 4);
    d. Results of serology (pos/neg, no titres) and nasopharyngeal swabs for PCR of HHCs at baseline, and at weeks 1, 2, 3 and 4);
2. Randomisation list
3. Qualitative and quantitative process data on intervention (STOPCOV database)
    a. Number and duration of calls to households
    b. transcripts from follow-up calls
    c. SMS/WhatsApp data
    d. transcripts from interviews with STOPCov counsellors and intervention/control index and HHC participants
4. Provincial health data(PID; all study participants and household contacts)

4. STUDY OUTCOMES

4.1. Primary study outcomes

- The incidence rate of SARS-CoV-2 infections among HHC of the index COVID19 cases over a 4 week period
- The time to transmission to HHC during that time period
- The association between the time to transmission and clinical symptoms of the index cases
- The symptomatic/asymptomatic ratio among transmitted SARS-CoV-2 in household contacts
- The proportion of transmissions to HHC in the enrolled households of COVID cases assigned to the infection mitigation intervention arm compared with the households in the enhanced usual care (EUC) arm.

4.2. Secondary Outcomes

- The proportion of index patients and HHC that visited a primary health clinic during the period of follow-up in the intervention arm as compared to the enhanced usual care (EUC) arm.
- The proportion of index patients and HHC that were hospitalised for COVID-19 during the period of follow-up in the intervention arm as compared to the enhanced usual care (EUC) arm.
- The proportion of index patients and HHC that died due to COVID-19 during the period of follow-up in the intervention arm as compared to the enhanced usual care (EUC) arm.
- The proportion of index patients and HHC that showed improved psychosocial functioning, defined as lower GHQ 12, UCL loneliness scores and Stigma scores, in the intervention arm as compared to the enhanced usual care (EUC) arm.
- The proportion of SARS-CoV-2 transmitted to HHC of the index COVID19 cases over a 4 week period, stratified by COVID-19 wave.
- The proportion of index patients and HHCs who developed antibodies during follow-up.
- Process Evaluation Outcomes:
  - Barriers and facilitators to implementation of the STOPCov intervention.
  - Identified contextual patterns within households that interact with the implementation of STOPCov intervention
  - Theoretical explanations of how different structural relations and mechanisms impact on intervention delivery and outcomes.
  - Generalizable inferences and predictions on how best to optimise the STOPCov intervention.

5. GENERAL PRINCIPLES OF ANALYSIS

The statistical analysis plan will be finalised and signed-off prior to database lock. Post-hoc analysis (analyses performed that are different from those described in the final version of the SAP) will be documented in the final report.

Descriptive summaries will be presented using summary statistics (e.g. number (n), mean, standard deviations (SD), median, minimum, maximum) and 95% confidence intervals (CI) for continuous parameters or frequency distributions (n, %) for categorical parameters.

In general, all households who entered the study will be included in the statistical evaluation (intention-to-treat analysis). Households who completed the final screening will be included in the per-protocol analysis. There will be no predefined interim statistical analysis or stopping rules.

All output will be generated using STATA SE Version 13 or R. No missing data will be replaced, and it will be shown in any summaries as missing.

6. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

6.1. Screening

The screening process will be described including presentation of numbers screened, numbers enrolled into the study and summary statistics on time between positive test of the index client and enrollment into the study.

6.2. Study Completion

The data on completion of the study and reasons for withdrawal will be summarized. The number of households or participants withdrawing and the reason for withdrawal will also be reported and summarized by time in the study.

6.3. Demography and Baseline Characteristics

Demographic data and baseline characteristics of index clients, households and household contacts will be summarized. Summary statistics will be adjusted for clustering by household.

6.4. Compliance

Uptake of the intervention will be described, including over time.

6.5. Protocol Deviations

All protocol deviations will be reported.

7. ANALYSIS OF PRIMARY STUDY OUTCOMES

7.1. The proportion of SARS-CoV-2 transmitted to HHC of the index COVID19 cases over a 4 week period

SARS-CoV-2 infection in a HHC will be defined as a positive PCR or evolving positive serologic test in a HHC who had a negative PCR or negative serology at baseline. The incidence of SARS-CoV-2 infection in HHC will be calculated as the number of new cases divided by the person time at risk and 95% confidence intervals calculated based on the Poisson distribution adjusted for clustering by household. This incidence as well as the timing to transmission will be analysed using exponential shared frailty models taking the clustering by household into account and allowing for censoring. The start time will be the date of diagnosis of the index case and the event time the date of diagnosis of the HHC using the date of the first positive test based on PCR or serology. Models will include possible confounders including adjustment for baseline values.

7.2. The proportion of transmissions to HHC in the enrolled households of index COVID cases assigned to the infection mitigation intervention arm compared with the households in the enhanced usual care (EUC) arm.

The incidence of SARS-Cov-2 will be estimated for the different study arms and incidence of and timing to SARS-CoV-2 per study arm will be compared by including study arm in the exponential frailty models.

8. ANALYSIS OF SECONDARY STUDY OUTCOMES

8.1. The proportion of index patients and HHC that were hospitalised for COVID-19 during the period of follow-up in the intervention arm as compared to the enhanced usual care (EUC) arm.

The binary outcome of hospitalisations for COVID-19 will be determined for the different study arms will be compared and analysed using a mixed effect logistic regression model.

8.2. The proportion of deaths in index patients and HHC due to COVID-19 during the period of follow-up in the intervention arm as compared to the enhanced usual care (EUC) arm.

The binary outcome of deaths in index patients and HHC for COVID-19 will be determined for the different study arms and will be compared and analysed using a mixed effect logistic regression model.

8.3. The proportion of index patients and HHC that showed improved psychosocial functioning in the intervention arm as compared to the enhanced usual care (EUC) arm.

The continuous outcome of psychosocial functioning in index patients and HHC for COVID-19 will be compared and analysed using linear regression.

8.4. The association between the time to transmission and clinical symptoms of the index cases

The incidence of transmitted SARS-CoV-2 infections to household contacts will be estimated and compared by the number and duration of clinical symptoms of the index patients. These symptom characteristics can also be included in the frailty models.

8.5. The symptomatic/asymptomatic ratio among transmitted SARS-CoV-2 in household contacts

The proportion of symptomatic and asymptomatic transmitted SARS-CoV-2 infections will be calculated, overall, with 95% confidence intervals based on binomial or asymptotic normal distributions.

8.6. The incidence of SARS-CoV-2 transmitted to HHC of the index COVID19 cases over a 4 week period, stratified by COVID-19 wave.

The incidence of SARS-CoV-2 will be compared by COVID-19 wave using exponential frailty models and including an indicator of the wave in the model.

The COVID19 wave will be defined as follows:

- wave 1: date of diagnosis of index case between 1 May and 31 November
- wave 2: date of diagnosis of index case from 1 December onwards

8.7. Analysis of household exposure to SARS-CoV-2 at baseline using the baseline serology result, and development of antibodies using the serology result during follow-up.

Longitudinal weekly prevalence profiles of serology results will be summarized overall and by study arm. These profiles can be analysed using hierarchical mixed effect logistic regression models with both household and HHC as nested random effects. Time to development of antibodies can be analysed using exponential frailty models. Models will include baseline serology results.

8.8. Process Evaluation Analysis

All interviews will be transcribed verbatim, and the analysis of qualitative data will be iterative, moving between data collection and data analysis to test emerging theories. It may for example emerge that some households have experienced significant stigma with their COVID-19 diagnosis, and this may shape their experience of the intervention, and require deeper exploration. The analysis of observational data will therefore require knowledge from the counsellors' interviews to compare how reported experiences relate to actual implementation and acceptability of the intervention. Care will be taken to identify and follow up 'deviant cases' which do not fit into emerging theories.

By setting the delivery of the STOPCov intervention within a macro, meso and micro contextual framework, we will be able to make the transition from the identification of patterns within households, to theoretical explanations of how different structural relations and mechanisms impact on intervention delivery and outcomes. Identifying 'telling cases' will facilitate generalizable inferences and predictions on how best to optimise the STOPCov intervention.