Statistical Analysis Plan

Evaluation of home-made cloth face mask on COVID-19 and respiratory illnesses prevention at the community level – a Cluster-Randomised Controlled Trial in urban Guinea-Bissau

NCT04471766

Bandim Health Project

21-09-2021

Contents

1. General analytical principles 2
1.1 Study population 2
1.2 Randomization 2
1.3 Follow-up period 2
1.3.1 Follow-up 2
1.3.2 Outcomes 3
1.3 Statistical model 3
1.4 Missing data 3
2. ANALYSES OF BASELINE COMPARABILITY 4
2.1 Baseline comparison 4
3. ANALYSES OF PRIMARY OUTCOMES 6
3.1 Self-reported COVID-19-like illness 6
3.2 Consultation for COVID-19-like illness 6
3.3 Severe illness 7
3.4 All-cause mortality 8
4. ADDITIONAL ANALYSES OF PRIMARY OUTCOMES 8
4.1 Sensitivity analyses of primary outcomes – intensity of follow-up 8
4.2 Sensitivity analyses of primary outcome – period of follow-up 8
4.3 Secondary analyses of primary outcome – Effect modifiers 8
4.4 Cause-specific death based on verbal autopsies 10
4.5 Changing effects during study period 11
REFERENCES 13
APPENDIX 14

1. General analytical principles

1.1 Study population

Eligible for participation were persons aged ≥10 years residing in one of the following three districts of Bandim Health Project's (BHP) urban Health and Demographic Surveillance System (HDSS): Cuntum II, Bandim I or Bandim II. There were no exclusion criteria.

The study population is the group of individuals fulfilling the enrolment criteria (age, residency) and who did not decline participation, and to whom masks were distributed in accordance with group allocation.

In some instances, there can be double registered individuals (persons who were registered twice in the BHP HDSS database). Such an individual will be excluded if included in both randomisation groups. If enrolled twice in the same group, the individual is considered from the first enrolment.

1.2 Randomization

Each of the three districts consist of numbered zones (Cuntum II: zones 2-4, Bandim I: zones 1-9, Bandim II: zones 1-8) and each zone is divided into subzones labelled by letters. All sub-zones were randomized to intervention or control stratified by study zone. To ensure balancing, 10,000 random combinations of subzones were drawn. Combinations with imbalances between the number of participants, age distribution above 50 years of age, and socioeconomic status (SES) (proportion of individuals from households with functioning electricity[1]) were dropped, and we selected at random between the remaining combinations (Appendix).

1.3 Follow-up period

1.3.1 Follow-up

Participants were followed up through telephone calls. We called all study participants after 4 months of follow-up. We had intended more frequent follow-up calls but were unable to implement that for all participants for logistic reasons. Thus, in a subset of the population (enrolled in some subzones of Cuntum II zone 2, 3 and 4 and in all subzones of Bandim I zone 8), we have implemented two telephone follow-up interviews: the 1st interview around 2 months of follow-up and the 2nd interview after 4 months of follow-up. In the remaining subzones, calls were only made after 4 months.

All participants also received a follow-up visit at home after 4 months of follow-up to ensure that we had complete information on vital status for all enrolled participants, and to distribute masks to participants in the control group.

[1] Source of data: most recent pregnancy between 2016-2019, household-level variable

1.3.2 Outcomes

Primary outcomes assessed:

- Self-reported COVID-19-like illness: Any episode of self-reported symptoms of COVID-19 (three or more of Constipation, Coughing, Fever, Dyspnea, Loss of sense of smell, Loss of sense of taste, Extreme fatigue).
- Consultation with COVID-19-like illness: Any health center or outpatient hospital consultation for COVID-19-like illness (as above) and/or positive COVID-19 test.
- Severe COVID-19-like illness: Any self-reported COVID-19-like illness (as above) and in-patient hospitalization or death.
- All-cause mortality: Death due to any cause during the follow-up period.

Secondary outcomes, which will be assessed:

- Cause-specific death
- Mask use compliance

Except for self-reported COVID-19-like illness (not dated, see 3.1), the main analyses of the primary outcomes will consider outcomes occurring during 4 months of follow-up, i.e., from the date of enrolment and until and including 121 days after enrolment. In sensitivity analyses, we will consider the entire period of observation, i.e., from date of enrolment to the time of interview (using the date of telephone call for information on consultation and hospitalization, and the date of home visit for information on mortality).

1.3 Statistical model

The proportion of individuals reporting a study outcome will be compared by group allocation in logistic regression models with generalised estimating equation (GEE) correction for subzone and adjusted for the factors used in the randomisation: number of people living in the subzone, proportion of individuals aged >50 years in the subzone and proportion of households with functioning electricity in each subzone. All statistical tests will be 2-tailed and $p \leq 0.05$ considered statistically significant.

1.4 Missing data

No data will be imputed. However, for a particular event, we will assume that it was not a "COVID-19-like illness" (see below) if no information is obtained. Some dates of events may be imprecise (e.g., unknown date of consultation or only month of hospitalization known) – such dates will be allocated to the midpoint of the range of possible dates. As all individuals are followed for mortality through the BHP HDSS and the home visits, individuals who we were not able to contact by telephone will still contribute to the assessment of effect on the mortality outcome.

2. ANALYSES OF BASELINE COMPARABILITY

2.1 Baseline comparison

We will describe participant flow by group allocation in a flowchart. Background factors will be summarised by counts (percentages), means (standard deviation) or medians (interquartile range) as appropriate by group allocation. Information on the proportion with missing information will be provided.

Box 1: Summary of background factors by intervention and control group, obtained either at enrolment ("E") or from the HDSS ("H")

- Sex **("E")**
- Age **("E")**
- District **("E")**
- Type of roof, veranda (flooring), ceiling, source of water supply (socio-economic factors (SES) **("E")**
- Functioning electricity (source of data: most recent pregnancy registered in the BHP HDSS between 2016-date of enrolment, family-level) (SES) **("H")**
- Number of people sleeping in the same room **("H")**
- Baseline willingness/intention (and when) to use the mask **("E")**
- For those refusing to wear mask: reason for refusal **("E")**

2.2. Information on follow-up visits

Originally, we intended to follow-up all participants at both 2 and 4 months, but logistically this was only possible for a subset of the population. At all follow-up interviews we inquired about mask use (used mask y/n, if yes - sometimes / most of the time / always when leaving the house) and exposure to large groups of people.

In all subzones, we also conducted direct observation from a distance on Monday-Friday from 8am-3pm (2hrs per "observation spot" within a defined group of houses) to register the number of people leaving the house and among those the number of people wearing a face mask. In addition to whether a mask was worn, we registered whether a worn mask covered the nose and mouth, and whether the used mask was the study cloth face mask. A total of three rounds of direct observation have been conducted (two full rounds and a third round to specific subzones).

Box 2: Summary of follow-up information by intervention and control group

The following information is collected at the follow-up interviews, and will be presented by randomization group:

Logistics:

- Number of follow-up telephone calls per individual
- Number of successful (yes vs no) telephone interviews

- Number of participants providing information him-/herself vs another household member providing information on their behalf
- Distribution of follow-up calls per interviewer (assistant)
- Time between enrolment and successful telephone follow-up call
- For home visits: Number of participants visited in household and with reported survival status; Time between enrolment and home visit

Exposure to large groups of people:

- Reported to have attended event with 20+ persons, who are not living with participant; (yes/no, and if yes, whether a frequent event (such as school, work or going to the market) vs occasional event (such as funeral, birthday or wedding celebration).
- Reported to have been in contact with a person with COVID-19

Mask compliance:

Mask use (used mask y/n, if yes - sometimes / most of the time / always when leaving the house)

- Self-reported mask use
    - Total
    - % Never (masc)
    - % Sometimes (mascquan)
    - % when leaving home (mascquan)
    - % Often (mascquan)

All information listed in this box will be presented by subset of population, who received a follow-up telephone call at 2 months vs at 4 months, allowing for a comparison of the quality of information obtained if there were two vs one telephone interview.

Apart from the direct participant contact, we will also compare mask compliance through the data obtained from direct observations in the study area:

- Total observed
- % using mask
- % using study mask
- % using mask covering nose and mouth

<u>The following analysis of mask compliance will be conducted for mask use (Y/N) and (always vs sometimes/never) and presented:</u>

```
xtset subzone
xtgee maskuse group c.sz_electricity c.sz_size  c.sz_prop>50, ///
family(binomial) link(logit) corr(exchangeable) robust eform
```

In similar models we will compare dichotomised measures of exposure and logistic factors by trial arm.

3. ANALYSES OF PRIMARY OUTCOMES

3.1 Self-reported COVID-19-like illness

Information collected on episodes of illness are not dated. Hence, the data analyses will focus on the full period elapsed between enrolment and telephone follow-up.

Table 1: Self-reported COVID-19-like illness

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures) according to randomization group.
End of follow- up	Telephone interview ≥4months after enrolment
Exclusions	- In intervention group, but no mask received - In control group, but received mask - No telephone follow-up performed - Double registered
Outcome	Any self-reported COVID-19-like illness (reported "yes" to three or more of the following symptoms: Constipation, Coughing, Fever, Dyspnea, Loss of sense of smell, Loss of sense of taste, Extreme fatigue) since enrolment at either 2- or 4-months follow-up call[2]. We will code this variable c19li as 0 for "yes to <3 of the listed symptoms" or 1 for "yes to ≥3 of the listed symptoms" for all individuals with an interview conducted.
Stata code	Analysis: xtset subzone xtgee c19li group c.sz_electricity c.sz_size c.sz_prop>50, family(binomial) /// link(logit) corr(exchangeable) robust eform (Where sz_electricity, sz_size and sz_prop>50 are the balancing variables (constant by subzone) used in the randomisation).

3.2 Consultation for COVID-19-like illness

Table 2: Consultation for COVID-19-like illness

[2]Self-reported means reported in telephone interview; can therefore be reported by participant or another household member living with participant (>18yrs).

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	Whichever comes first: the last telephone interview or 121 days after enrolment
Exclusions	- In intervention group, but no mask received - In control group, but received mask - No telephone follow-up performed - Double registered
Outcome	Any reported consultation for COVID-19-like illness within 121 days after enrolment: g consult = [(consultation date>date enrolment, consultation date ≤ date enrolment+121) and (c19li==1 or positive COVID-19 test)].
Stata code	Analysis: xtset subzone xtgee consult group c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform

3.3 Severe illness

Table 3: Severe COVID-19-like illness

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	121 days after enrolment
Exclusions	- In intervention group but no mask received - In control group but received mask - No telephone follow-up performed - Double registered
Outcome	Any reported hospitalization for or death from COVID-19-like illness within 121 days after enrolment: g severe = [((hospitalization date>date enrolment, hospitalization date ≤ date enrolment+121) and (c19li==1)) or ((death date>date enrolment, death date ≤ date enrolment+121) and (c19li==1))].
Stata code	Analysis: xtset subzone xtgee severe group c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform

3.4 All-cause mortality

Table 4: All-cause mortality

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	121 days after enrolment
Exclusions	- In intervention group but no mask received - In control group but received mask - Double registered
Outcome	Any reported death due to any cause within 121 days after enrolment: death = [(death date>date enrolment, death date ≤ date enrolment+121)]. In the case of divergent information between telephone interview and household visit, the information given at household visit is considered superior.
Stata code	xtset subzone xtgee death group c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform

4. ADDITIONAL ANALYSES OF PRIMARY OUTCOMES

4.1 Sensitivity analyses of primary outcomes – intensity of follow-up

To assess whether information on morbidity, and thus the completeness of data, differ for the subset of the population followed closer (at both 2 and 4 months) and those only interviewed by 4 months, we will compare the incidence of morbidity for participants, who received two telephone follow-up interviews (at both 2 and 4 months after enrolment) with the incidence of morbidity among those who received only one follow-up call. Furthermore, we will investigate if results differ according to intensity of follow-up. This will be analysed by including an 'intensity of follow-up'-variable as a potential effect modifier (see table 5 for analyses syntax for potential effect modifiers).

4.2 Sensitivity analyses of primary outcome – period of follow-up

In sensitivity analyses, we will assess if extending the period of follow-up to the date of the telephone interview/home visit affects the conclusions.

4.3 Secondary analyses of primary outcome – Effect modifiers

Secondary analyses of the primary outcome aim to assess if the effect of the mask distribution varies with potential effect modifiers identified as important for the severity of COVID-19 or for the effect of masks in prior studies.

In analyses of the primary outcome, we will assess the effect of the following potential effect modifiers allowing the effect of the intervention to vary with:

- Sex: Studies have indicated a higher COVID-19 case fatality in men compared with women[1, 2].
- Age: A recent trial from Bangladesh identified age as a potential effect modifier of the effects of mask use to prevent COVID-19[3].
- Factors interpreted as indicators of higher risk of exposure:
    - Having a ceiling (Y/N): Many houses are multifamily houses with open air circulation under the roof. A ceiling would be expected to limit the circulation of respiratory pathogens to other families under the same roof (collected at baseline).
    - Attending events with many people defined as an event with 20+ attending persons, who are not living with the individual (collected at follow-up).
    - Reported to have been in contact with a person with COVID-19 (collected at follow-up).
    - Number of children <10 registered in the household on the day of enrolment.
    - Number of persons sleeping in the same room registered on the day of enrolment.
    - For morbidity outcomes (COVID-19-like illness and consultation) only: intensity of follow-up.

The individual records will remain as in the primary analysis, and the model allow for interaction with the potential effect modifier.

Table 5: Effect modifiers

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	121 days after enrolment
Exclusions	- In intervention group, but no mask received - In control group, but received mask - No telephone follow-up performed - Double registered
Outcome	Self-reported COVID-19-like illness, Consultation, Severe illness, Mortality (as specified in tables above).
Effect modifiers (EfM) in 5 separate models)	Sex: (M/F) Age (main analyses: Age above vs below 50 years and decade: 10-19, 20-29, etc.) Reported to have attended event(s) with 20+ people not living with participant (Y/N) Reported to have been in contact with a person with COVID-19 (Y/N) Ceiling (Y/N) Number of children <10 years of age registered in the household on the day of enrolment (source of data: BHP HDSS) Number of persons sleeping in the same room (source of data: BHP HDSS) Morbidity outcomes only: Intensity of follow (follow-up at 2- and 4-months vs 4 months only)

Stata code	Analysis: xtset subzone xtgee outcome group#EfM EfM c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform contrast group#EfM (where EfM= sex / age / ceiling / attended events / assisted covid / <10children / no sleep / intensity of follow-up).

4.4 Cause-specific death based on verbal autopsies

Verbal autopsies will be collected for all deaths among participants to determine the cause of death for all deaths registered during the study period. Using the same statistical approach as in the primary analysis, we will assess whether the overall effect of distributing face masks differ by cause of death based on classifications as likely COVID-19, other infectious, non-infectious.

Note: cause of death will only be available later, after verbal autopsies have been conducted, and for that reason may not be included in the first publication based on this trial.

Table 6: Cause-specific death

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	121 days after enrolment
Exclusions	- In intervention group, but no mask received - In control group, but received mask - Double registered
Outcome	Any reported death due to COVID19, other infectious, non-infectious cause within 121 after enrolment: Cause1-3 = [(death date>date enrolment, death date ≤ date enrolment+121) and (cause=COVID19/Infectious/noninfectious)]. (where death is coded 1 for a death within 4 months of follow-up, and cause of death is coded as likely COVID-19, other infectious, non-infectious based on the information collected through the verbal autopsies).
Stata code	Analysis: xtset subzone xtgee cause1-3 group c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform

4.5 Changing effects during study period

During 2020 and 2021, Guinea-Bissau experienced three waves of COVD-19 infections, during which the risk of COVID-19 has been elevated (figure). Based on this, we have identified periods of higher transmission: Before 1 October 2020, 1 January - 1 April 2021 and after 1 July 2021.

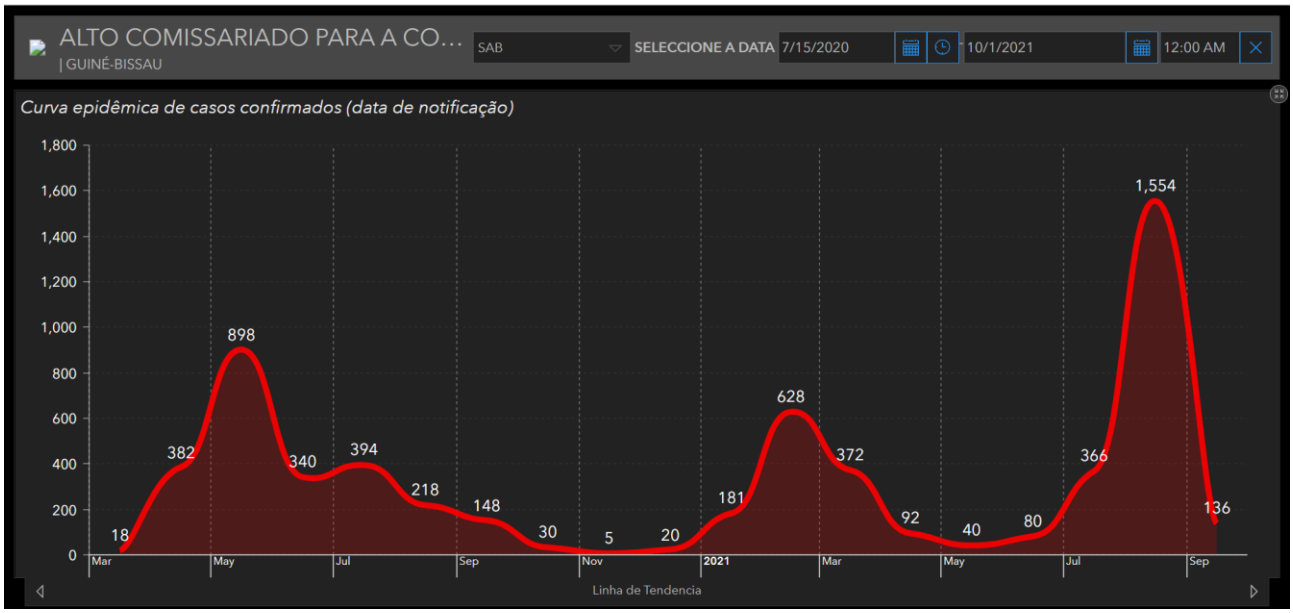

Figure: Number of confirmed COVID-19 cases per week in the Bissau sector (source www.ACCOVID.com).

We will generate a continuous measure of the proportion of follow-up time within these high exposure periods and assess whether the effect of the intervention varies by randomisation group.

Table 7: Potential effect modifiers of the primary outcomes - periods of exposure to waves of Covid-19 (high exposure)

Population	Randomised participants receiving two local-made masks, mask instruction and information on COVID-19 preventive measures (vs only information on COVID-19 preventive measures).
End of follow-up	121 days after enrolment
Exclusions	- In intervention group, but no mask received - In control group, but received mask - No telephone follow-up performed - Double registered -
Outcome	Self-reported COVID-19-like illness, Consultation, Severe illness, Mortality (as specified in tables above).
Stata code	Analysis: xtset subzone

	xtgee outcome group#c.wave c.wave c.sz_electricity c.sz_size c.sz_prop>50, /// family(binomial) link(logit) corr(exchangeable) robust eform contrast group#c.wave (where wave is a continuous variable from 0-1 indicating the proportion of follow-up time as highly exposed).

REFERENCES

1. Bhopal SS, Bhopal R. Sex differential in COVID-19 mortality varies markedly by age. The Lancet. 2020;396(10250):532-3.
2. Gebhard C, Regitz-Zagrosek V, Neuhauser HK, Morgan R, Klein SL. Impact of sex and gender on COVID-19 outcomes in Europe. Biol Sex Differ. 2020;11(1):29-.
3. Abaluck J, Kwong LH, Styczynski A, Haque A, Kabir A, Bates-Jeffries E, et al. The Impact of Community Masking on COVID-19: A Cluster-Randomized Trial in Bangladesh. 2021.

APPENDIX

Constrained randomisation procedure for the MASK project

Constrained randomisation: the study will block-randomise 234 sub-zones within six urban district of the capital Bissau of Guinea-Bissau to either received the intervention (masks) or not (control) in a 1:1 ratio. To ensure that key covariates are balanced within the randomization groups, we will use constrained randomization on the following covariates: distribution of total number of participants, proportion of participants aged 50 years or above and the proportion of households with functioning electricity.

The covariates (raw data):

Total number of expected participants is based on total number of people registered in the census at the Bandim Health Project on the 1st of January 2015. This number is adjusted with the expected refusal of 10%.

Number of expected participants aged 50 years and above is based on total number of people registered in the census at the Bandim Health Project on the 1st of January 2015

Number of household with functioning electricity is based on all pregnancies registered between 2016-2019 and the information collected for the associated pregnancy termination interview. Only the information obtained from the last pregnancy from each household (g household_id = tabz_string + housegrp + camo_string + fam_string) was used

The setup: The urban district has a total of 234 sub-zones in the six districts, ranging from 19 (Mindara) to 94 (Bandim 1). Effectively this means that the number of hypothetical valid allocations ranges between 184 756 and $8.1 \cdot 10^{26}$, which practically makes it impossible to check all potential allocations as described in Moulton 2004.

With the almost endless hypothetical allocations, to ensure that we obtain a balanced allocation of the constrained covariates for the project, we draw 10 000 random potential allocations (for each district) and use the 2.5% and 97.5% percentiles as lower and upper cut-offs for acceptable values for the constrained covariates. Thus, we remove the most extreme cases of allocation and arrive at distributions, which vary between:

> Absolute number of persons / arm – distribution varies from balanced 50/50 to 48%/52%
>
> Age>50: Difference in proportion of population among individuals >10 years, who are also above 50 years (based on data 2015) +/2% (median ca 10%)
>
> Proportion of individuals with electricity in the household balanced between groups (difference 6-11%)

Algorithm for drawing the allocation (randomisation):

1) Set a fixed seed before doing the allocation.
2) Draw an allocation for the first district.
3) Check whether the distributions of all three covariates are within the simulated cut-offs.
4) If one or more of the covariates fail step 3, repeat step 2-3 until it passes.
5) Do step 2-4 for the last five districts.
6) You now have six vectors containing the balanced allocation of intervention and control for the constrained variables.