Statistical Analysis Plan

BCG-CORONA

Title:	REDUCING HEALTH CARE WORKERS ABSENTEEISM IN SARS-CoV-2 PANDEMIC BY ENHANCED TRAINED IMMUNE RESPONSES THROUGH BACILLUS CALMETTE-GUÉRIN VACCINATION, A RANDOMIZED CONTROLLED TRIAL
Sponsor:	University Medical Center Utrecht (UMCU)
SAP version - date:	V2.0 of 02 May 2020

SAP authors:

Sponsor's representative, Coordinating principle investigator	Marc J.M. Bonten, MD, PhD Dpt. of Medical Microbiology University Medical Center Utrecht Heidelberglaan 100, 3584 CX Utrecht, the Netherlands Tel: +31-75-57676 E-mail: m.j.m.bonten@umcutrecht.nl
Coordinating investigator	Thijs ten Doesschate, MD Julius Center for Health Sciences and Primary Care University Medical Center Utrecht Heidelberglaan 100, 3584 CX Utrecht, the Netherlands Tel: +31-88-75-69236 E-mail: T.tenDoesschate@umcutrecht.nl
Trial statistician	Cornelis H. van Werkhoven, MD, PhD Julius Center for Health Sciences and Primary Care University Medical Center Utrecht Heidelberglaan 100, 3584 CX Utrecht, the Netherlands Tel: +31-88-75-68171 E-mail: c.h.vanwerkhoven@umcutrecht.nl

Signature page

Title: **REDUCING HEALTH CARE WORKERS ABSENTEEISM IN SARS-CoV-2 PANDEMIC BY ENHANCED TRAINED IMMUNE RESPONSES THROUGH BACILLUS CALMETTE-GUÉRIN VACCINATION, A RANDOMIZED CONTROLLED TRIAL**

Document release - date: Statistical analysis plan V2.0 of 02 May 2020

Approval page

Marc. J.M. Bonten, MD, PhD **Coordinating principle investigator** Dpt. of Medical Microbiology University Medical Center Utrecht Date: 07.05.2020 Signature: 	**Cornelis H. van Werkhoven, MD, PhD** **Trial statistician** Julius Center for Health Sciences and Primary Care University Medical Center Utrecht Date: 02 May 2020 Signature: 

Table of contents

List of abbreviations

BCG:	Bacillus Calmette-Guérin
CI:	Credible Interval
eCRF:	electronic Case Report Form
ICU:	Intensive Care Unit
IQR:	Inter Quartile Range
RCT:	Randomised Controlled Trial
SD:	Standard deviation
UMCU:	University Medical Center Utrecht

1 Overview of the study

This is a copy of the protocol summary, version 3.0, 01-April-2020:

Rationale: SARS-CoV-2 spreads rapidly throughout the world. A large epidemic in the Netherlands would seriously challenge the available hospital capacity, and this would be augmented by absenteeism of healthcare workers (HCW). Strategies to prevent absenteeism of HCW are, therefore, desperately needed to safeguard continuous patient care. Bacille Calmette-Guérin (BCG) is a vaccine against tuberculosis, with protective non-specific effects against other respiratory tract infections in *in vitro* and *in vivo* studies, and reported significant reductions in morbidity and mortality. We hypothesize that BCG vaccination can reduce HCW absenteeism during the epidemic phase of SARS-CoV-2.

Objective: Primary objective: To reduce absenteeism among HCW with direct patient contacts during the epidemic phase of COVID19. Secondary objective: To reduce hospital admission, ICU admission or death in HCW with direct patient contacts during the epidemic phase of COVID19.

Study design: A placebo-controlled adaptive multi-centre randomized controlled trial.

Study population: HCW with direct patient contacts, defined as nurses and physicians working at emergency rooms and wards where COVID-infected patients are treated.

Intervention: Participants will be randomized between intracutaneous administration of BCG vaccine or placebo in a 1:1 ratio.

Main study parameters/endpoints: Primary endpoint: number of days of (unplanned) absenteeism for any reason. Secondary endpoints: number of days of (unplanned) absenteeism because of documented SARS-CoV-2 infection, and the cumulative incidence of hospital admission, Intensive Care Admission, and death.

Nature and extent of the burden and risks associated with participation, benefit and group relatedness: Based on previous experience and randomized controlled trials in adult and elderly individuals, the risks of BCG vaccination are considered low. The objective of this trial is to evaluate the beneficial effects of BCG vaccination through a lower work absenteeism rate of HCW and/or a mitigated clinical course of SARS-CoV-2 infection. The primary endpoint and the adaptive design with frequent interim analyses facilitate maximum efficiency of the trial, so that results can inform policy making during the ongoing epidemic.

2 Sample size determination

The aim is to enrol a total of 1500 HCW (750 BCG and 750 placebo). For a detailed sample size calculation, please refer to the protocol.

3 Definitions

3.1 Treatment groups

BCG: participants randomised to receive BCG

Placebo: participants randomised to receive placebo

3.2 Analysed populations

The primary anlysis population is the intention-to-treat (ITT) population in which participants will be analysed according to the random allocation.

Subgroups:

No subgroups have been predefined.

3.3 Study definitions

COVID-19 disease is defined as respiratory tract symptoms or fever in combination with PCR-based detection of SARS-CoV-2 in a respiratory sample.

SARS-CoV-2: severe acute respiratory syndrome coronavirus-2. The virus that causes COVID-19 disease.

4 General statistical considerations

4.1 General statistical methodology

The methodology presented in this statistical analysis plan takes into account methods planned in the protocol, specifies them, and completes them. The analysis will be performed in the statistical software R.[1] The version of R and R libraries used will be reported.

4.2 Handling of missing data

For analyses of primary or secondary endpoints with missing data, multiple imputed data analysis will be performed using Multivariate Imputation by Chained Equations; complete data analysis will be performed as a sensitivity analysis to check for inconsistencies.[2] Missing outcome data will not be imputed.

4.3 Descriptive statistics

According to the variable type, descriptive statistics will be:

- Quantitative criteria: number of observations (N), mean, standard deviation (SD), median and inter-quartile range (IQR), as applicable.
- Qualitative criteria: number of observations (N), absolute frequency (n) and relative frequency (%). Percentages will be calculated on the number of participants with documented data.

4.4 Inferential statistics

Bayesian analysis will be used for inferential statistics. Effect estimates for primary and secondary endpoints will be reported with 95% credible intervals (CI). For the primary endpoint, a 99% CI as well as the posterior probability of superiority and futility will be reported in addition to the 95% CI, in line with the interim analysis.

5 Disposition of participants and conduct of the study

5.1 Participant disposition and size of the analysed populations

The study flowchart will describe:

- Number of subjects screened
- Number of subjects not meeting the eligibility criteria
    - Reasons for non-eligibility
- Number of subjects eligible but not included
    - Reasons for non-inclusion
- Number of participants included and randomised
- Early termination: Number of participants withdrawn, dropouts
    - Reasons for early termination

5.2 Conduct of the study

The following will be reported on the conduct of the study:

- Enrolment period: date First Patient First Visit – date Last Patient First Visit
- Distribution of actual duration of follow-up per participant by treatment arm
- Number and proportion of days with missing outcome measurement by treatment arm

6 Demographics and other baseline characteristics

6.1 Demographics / baseline characteristics

The following data will be described for included participants by treatment arm and for the total population:

VARIABLE	DERIVATION	STATISTICAL METHOD *
Age (years)	Self-reported; baseline eCRF	Median (IQR) or mean (sd) [1]
Gender	Self-reported; baseline eCRF: Male/ Female	n/N (%)
History of BCG vaccination	Self-reported; baseline eCRF: Yes/No	n/N (%)
Age at BCG vaccination	Self-reported; baseline eCRF	Median (IQR) or mean (sd) [1] among those with reported history of BCG vaccination
History of positive Mantoux test	Self-reported; baseline eCRF. Options are: - Yes - No; at least one negative test - No; never tested Answer option "I don't know" will be handled as missing value	n/N (%) for all categories
History of positive TB quantiferon test	Self-reported; baseline eCRF. Options are: - Yes - No; at least one negative test - No; never tested Answer option "I don't know" will be handled as missing value	n/N (%) for all categories
Receipt of influenza vaccination in past season	Self-reported; baseline eCRF: Yes/No	n/N (%)
Receipt of other vaccinations in past 12 months	Self-reported; baseline eCRF (open question) [2]	n/N (%) for each reported vaccination
History of respiratory tract infection in past winter	Self-reported; baseline eCRF. Options are: - Yes, with fever - Yes, without fever - No	n/N (%) for all categories:
Use of anti-hypertensive medication	Self-reported; baseline eCRF: Yes/No	n/N (%)
History of cardiovascular disease	Self-reported; baseline eCRF: Yes/No	n/N (%)
Use of anti-diabetic medication	Self-reported; baseline eCRF: Yes/No	n/N (%)
History of asthma	Self-reported; baseline eCRF: Yes/No	n/N (%)
History of other pulmonary diseases	Self-reported; baseline eCRF (open question) [2]	n/N (%) for each reported pulmonary disease

History of smoking	Self-reported; baseline eCRF. Options are: - Never smoked - Ever smoked - Current smoking	n/N (%) for all categories
History of allergic rhinitis	Self-reported; baseline eCRF: Yes/No	n/N (%)
Number of household members	Self-reported; baseline eCRF	Median (IQR) or mean (sd) [1]
Age of household members	Self-reported; baseline eCRF	n/N (%) with household members < 18 n/N (%) with household members >= 18 and < 50 n/N (%) with household members >= 50
Department of employment	Self-reported; baseline eCRF. Options are: - Intensive care - Medium care - Emergency room - Internal medicine - Infectious diseases - Pulmonary diseases - Other, specify (open question)[2]	n/N (%) for all categories
Position at department of employment	Self-reported; baseline eCRF. Options are: - Nurse - Medical doctor - Paramedical - Supportive - Secretary	n/N (%) for all categories
Current or planned deployment in dedicated COVID-19 department	Self-reported; baseline eCRF. Yes/No Answer option "I don't know" will be handled as missing value	n/N (%)
Average number of working days in the hospital	Self-reported; baseline eCRF.	Median (IQR) or mean (sd) [1]
Average number of working days on evening/night shift	Self-reported; baseline eCRF.	Median (IQR) or mean (sd) [1]
Percentage of working time spend with direct patient contact	Self-reported; baseline eCRF. Options are: - ≤25% - 26-50% - 51-75% - >75%	n/N (%) for all categories
Has at least one day of work loss due to illness between 1 January and 15	Self-reported; baseline eCRF. Options are:	n/N (%) for all categories

March 2020	- Yes, due to respiratory tract infection - Yes, for alternative health issue - No	
Total days of work loss due to illness between 1 January and 15 March 2020	Self-reported; baseline eCRF.	Median (IQR) or mean (sd) [1] among those with reported work loss due to illness
Previously tested for COVID-19	Self-reported; baseline eCRF. Options are: - No - Yes, negative test result - Yes, positive test result	n/N (%) for all categories
Number of days since negative test	Self-reported; baseline eCRF. Date negative test minus date of enrolment.	Median (IQR) or mean (sd) [1] among those with reported negative test result
Number of days since positive test	Self-reported; baseline eCRF. Date positive test minus date of enrolment.	Median (IQR) or mean (sd) [1] among those with reported positive test result
Has had negative COVID-19 test after being tested positive	Self-reported; baseline eCRF. Yes / No	n/N (%)
Number of days since negative test after being tested positive	Self-reported; baseline eCRF. Date negative test minus date of enrolment.	Median (IQR) or mean (sd) [1] among those with reported negative test result after being tested positive

* It will not be statistically tested whether baseline characteristics differ by treatment arm.

1. The description method will depend on the normality of the data.

2. Open questions will be categorized by the coordinating investigator.

7 Analysis of the primary endpoint

The primary endpoint, work absenteeism for any reason, will be reported as the average proportion of sick-days with standard deviation by treatment arm. It will be analysed as total counts (i.e. one observation per participant) using a Bayesian negative binomial regression with a fixed effect for BCG, hospital, enrolment week (categorical), age, department (whether planned to work on COVID-19 dedicated department), sick leave prior to enrolment (as a proportion of fte), and presence of at least one of four comorbidities: cardiovascular disease, use of anti-diabetic medication, asthma, pulmonary diseases. Age and prior sick leave will be modelled using a spline function. The total number of planned workdays for the respective person over the follow-up period will be used as offset (log-transformed). The brm function from the R package brms will be used for fitting the negative binomial model.[3]

The effect will be reported as a relative risk with 95% and 99% CI. The posterior probability for the superiority hypothesis (RR < 1) will be reported as well as the posterior probability for the futility hypothesis (RR > 0.8).

Model assumptions will be checked using residuals plots and rootograms and posterior distribution plots will be presented. If model assumptions are violated, the following alternative model will be used: analysis as sick/non-sick working days (i.e. multiple observation per participant, one observation for each planned working day) using a Bayesian logistic regression model with a random intercept per participant, a fixed effect for BCG, hospital, observation week (categorical), age, department (whether planned to work on COVID-19 dedicated department), sick leave prior to enrolment (as a proportion of fte), and presence of at least one of four comorbidities: cardiovascular disease, use of anti-diabetic medication, asthma, other pulmonary diseases. Age and prior sick leave will be modelled using a spline function. An autocorrelation structure of 1st order over time within participants will be added. This model yields an odds ratio (OR) which can be interpreted as RR if the risk of sick leave is <10% and will be used for determination of superiority or futility in the same manner as planned for the RR.

The same modelling approach will be used during interim analyses. Please refer to section 10: Interim analysis for a description.

8 Analysis of the secondary endpoints

8.1 number of days of unplanned absenteeism, because of COVID-19 disease

The same modelling approach will be used as for the primary endpoint.

8.2 Cumulative incidence of COVID-19 disease

This will be analysed as time-to-event outcome. A proportional hazards model will be used, adjusted for hospital, enrolment week (categorical), age, department (whether planned to work on COVID-19 dedicated department), sick leave prior to enrolment (as a proportion of fte), and presence of at least one of four comorbidities: cardiovascular disease, use of anti-diabetic medication, asthma, other pulmonary diseases. Age and prior sick leave will be modelled using a spline function. When applicable (i.e. in the event that one or more participants have died during the follow-up period) a competing events analysis will be performed in addition (Fine & Gray model).

8.3 Cumulative incidence of self-reported acute respiratory symptoms or fever

Same analysis as 8.2 Cumulative incidence of COVID-19 disease.

8.4 Cumulative incidence of hospital admission due to COVID-19 disease

Same analysis as 8.2 Documented infection with SARS-CoV-2.

8.5 Cumulative incidence of intensive care admission due to COVID-19 disease

Same analysis as 8.2 Documented infection with SARS-CoV-2.

8.6 Cumulative incidence of death due to COVID-19 disease

Same analysis as 8.2 Documented infection with SARS-CoV-2. If deaths not due to documented SARS-CoV-2 infection have occurred, this will be considered a competing event.

9 Analysis of the exploratory endpoints

9.1 Safety endpoints: (Serious) adverse events (AE / SAE) and suspected unexpected serious adverse reactions (SUSAR)

Frequencies of AEs, SAEs and SUSARs will be described per intervention arm for total number of events, per solicited event, and for unsolicited events. Frequencies of having at least one AE / SAE / SUSAR will be compared using logistic regression analysis and reported as OR with 95% CI.

For the following endpoints, the same modelling approach will be used as for endpoint 8.2 Cumulative incidence of COVID-19 disease:

9.1.1 the cumulative incidence of hospital admission for any reason

9.1.2 the cumulative incidence of intensive care admission for any reason

9.1.3 the cumulative incidence of death for any reason

9.2 Absenteeism endpoints

The same modelling approach will be used as for the primary endpoint.

9.2.1 number of days of absenteeism, because of imposed quarantine as a result of exposure to SARS-CoV-2 infection

9.2.2 number of days of absenteeism, because of imposed quarantine as a result of having acute respiratory symptoms, fever or documented SARS-CoV-2 infection

9.2.3 number of days of unplanned absenteeism because of self-reported acute respiratory symptoms

9.3 Days with symptoms endpoints

For the following endpoints, the same modelling approach will be used as for the primary endpoint.

9.3.1 number of days of self-reported fever (≥38 °C)

9.3.2 number of days of self-reported acute respiratory symptoms

For the following endpoints, the same modelling approach will be used as for endpoint 8.2 Cumulative incidence of COVID-19 disease

9.3.3 the cumulative incidence of self-reported fever (≥38 °C)

9.3.4 the cumulative incidence of self-reported acute respiratory symptoms

9.4 Immunological endpoints

Analysis of immunological endpoints are not pre-specified and will be described in the publication reporting on these endpoints. This related to the following:

9.4.1 the incidence and magnitude of plasma/serum antibodies (IgA,M,G) and SARS-CoV-2-specific antibodies at the end of the study period

9.4.2 the incidence and magnitude of total and/or SARS-COV-2-specific mucosal antibodies at the end of the study period

10 Interim analyses

Interim analyses will be performed starting from 4 weeks after the first enrolled patient and subsequently every 2 weeks. The interim analysis will be performed by the trial statistician unblinded and will not be disclosed to other study team members. Monthly, the results will be provided to the DSMB unblinded. In case of suggested futility or efficacy, the DSMB statistician will independently replicate the full data analysis before drawing conclusions.

The disposition of participants and conduct of the study will be analysed and reported (section 5).

Of the baseline characteristics (section 6), the following will be reported during the interim analysis:

- Hospital of employment
- Age
- Gender
- History of BCG vaccination
- Department of employment
- Whether planned to work on COVID-19 dedicated department
- Position at department of employment
- Average number of working days in the hospital
- Total days of work loss due to illness between 1 January and 15 March 2020

The following endpoints will be analysed and reported in the same way as described in the above sections:

- Primary endpoint, work absenteeism for any reason (section 7). If during any of the interim analyses, the posterior probability of superiority is > 0.995 or the posterior probability of futility is > 0.99, a conclusion is reached. In case of superiority, the trial will be stopped. In case of futility, the DSMB will take the totality of the data into account in their recommendation to continue or stop the trial.
- Hospital admission due to COVID-19 (section 8.13)
- SAEs and SUSARs (section 9.1)

Please also refer to the DSMB charter for roles and responsibilities and organizational aspects.

11 References

1. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
2. Stef van Buuren, Karin Groothuis-Oudshoorn (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. URL https://www.jstatsoft.org/v45/i03/.
3. Paul-Christian Bürkner (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 1-28. doi:10.18637/jss.v080.i01