COG-UK HOCI

COG-UK Project Hospital-Onset COVID-19 Infections Study

A phase III prospective, interventional, cohort, superiority study to evaluate the benefit of rapid COVID-19 genomic sequencing (the COVID-19 GENOMICS UK project) on infection control in preventing the spread of the virus in United Kingdom NHS hospitals

Protocol version: V5.0 dated 18.03.2021
IRAS ID: 283014

STATISTICAL ANALYSIS PLAN (SAP)

VERSION 1.0, 21ST APRIL 2021

Authorised by:	**Signature**	**Date**
Prof Judith Breuer, Chief Investigator Email: j.breuer@ucl.ac.uk ☐ I Approve this document	DocuSigned by: Judith Breuer D72BC4B9BFFA481...	21-Apr-2021
Dr Andrew Copas, Trial Statistician Email: a.copas@ucl.ac.uk ☐ I Approve this document	DocuSigned by: Andrew Copas A39B1D40B0AF463...	21-Apr-2021

First draft prepared by:
Oliver Stirrup, Statistician
Email: oliver.stirrup@ucl.ac.uk

Table of contents

1 ABBREVIATIONS

Acronyms	Meaning
CCTU	Comprehensive Clinical Trials Unit
COG-UK	COVID-19 Genomics UK Consortium
COVID-19	Coronavirus disease 2019
CRF	Case Record Form
ED	Emergency department
GP	General practitioner
HAI	Hospital-acquired infection
HCAI	Health care-associated infection
HOCI	Hospital onset COVID-19 infection
HCW	Health care worker
IPC	Infection Prevention and Control
ITT	Intention to Treat
LRT	Likelihood ratio test
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SNP	Single nucleotide polymorphism
SRT	Sequence reporting tool
TMF	Trial Master File
TMG	Trial Management Group
TSC-DMC	Trial steering committee – data monitoring committee
UCL	University College London

2 DEFINITIONS

HCAI: Effectively another term for a hospital acquired infection (HAI), but specifically used by IPC teams when classifying HOCI cases according to likelihood of nosocomial (i.e. within hospital) infection based on interval from admission to positive test: 3-7 days post admission = indeterminate HCAI; 8-14 days post admission = probable HCAI; >14 days post admission = definite HCAI.

HOCI: An inpatient with SARS-CoV-2 infection with positive sample >48 hours from admission and without suspicion of COVID-19 at time of admission, representing a possible nosocomial infection.

IPC-defined SARS-CoV-2 HAI: Any HOCI case with an interval of ≥8 days from admission to symptom onset (if known) or sample date (i.e. those meeting the definition of a probable or definite HCAI).

IPC-defined SARS-CoV-2 hospital outbreak: At least two HOCI cases on the same ward, with at least one having an interval of ≥8 days from admission to symptom onset (if known) or sample date, and with the outbreak event considered to be concluded if there is a gap of 28 days before the observation of another HOCI case.

IPC+sequencing-defined SARS-CoV-2 hospital outbreak: At least two HOCI cases on the same ward that form a genetic cluster with maximum viral sequence pairwise SNP distance of 2 between each individual included and their nearest neighbour within the cluster.

3 BACKGROUND AND DESIGN

Aim and objectives

The aim of the study is to determine the effectiveness of sequencing SARS-CoV-2 genomes in order to inform infection prevention and control team (IPC) efforts and to reduce the incidence of healthcare associated infections (HCAIs). The study has a sequential phase design with intervention periods defined at site level, with each site having an initial baseline period with standard of care or no sequencing, and intervention phases with slow or rapid sequencing of SARS-CoV-2 cases. Data will be gathered on cases of hospital onset COVID-19 infection (HOCIs), defined as those with interval from admission to first positive test result >48 h without suspicion of COVID-19 at admission, and a summary report on sequencing results will be returned to IPC teams on each such case during the intervention periods.

This SAP specifies the study results to be included in the primary analysis publication, and does not cover the health economics or process evaluation analyses described in the overall study Protocol.

The objectives of the trial, as stated in the Protocol, are:

- Primary: To determine whether the use of rapid and/or slow sequencing reduces the incidence of SARS-CoV-2 HCAIs in hospital settings.

- Primary: During intervention phases, to evaluate whether the use of rapid and/or slow sequencing identifies nosocomial (i.e. within hospital) transmission where this was not identified by pre-sequencing IPC evaluation.

- Secondary: To determine whether the use of rapid and/or slow sequencing reduces the incidence of IPC-defined SARS-CoV-2 outbreak events in hospital settings.

- Secondary: To evaluate the incidence of ‘IPC+ sequencing’-defined SARS-CoV-2 outbreak events in hospital settings during study phases with reporting of sequencing results.

- Secondary: To evaluate changes to IPC actions implemented following receipt of SARS-CoV-2 sequence report for a given HOCI.

- Secondary: To evaluate changes to IPC actions that would ideally have been implemented following receipt of SARS-CoV-2 sequence report for a given HOCI.

- Secondary: To evaluate staff absence through illness on those wards impacted by HOCI cases.

Definitions regarding interpretation of these objectives into specific outcome measures are given late in the SAP.

Population studied

Study sites are all general hospitals. Data are recorded in all phases for all patients meeting the following inclusion criteria:

- Participants must have confirmed SARS-CoV-2 infection and be a hospital-onset COVID-19 infection (HOCI) with first confirmed test for SARS-CoV-2 >48 hours after admission and without suspicion of COVID-19 at time of admission.

- Participants must have provided a saliva sample or a nasal swab/pharyngeal swab / combined nasal and pharyngeal swab / nasopharyngeal aspirate or broncho alveolar lavage sample for evaluation in the COG-UK project.

- Participants may be of any age to be included in study

There are no exclusion criteria for recording of patient data.

Study design
This study is a sequential trial of baseline and intervention conditions at the site-level, without randomisation of phase order. Data on HOCI cases are collected during all study phases, and IPC conclusions and actions informed by the sequencing report are recorded during the intervention phases.

Consent for participation in the study
Consent for patient involvement will not be sought for the COG-UK HOCI study. This approach relies on the Health Service (Control of Patient Information) Regulations 2002 (SI 1438).

Blinding:
The conduct of the study itself does not involve blinding of healthcare professionals or statisticians, as this would not be possible given the nature of the intervention and the sequential study design. Individual patients for whom data are collected are not directly affected by the intervention, and so patient blinding is not relevant.

4 OUTCOME MEASURES

4.1 Primary outcomes

- Weekly incidence of IPC-defined SARS-CoV-2 HAIs, measured as incidence rate of recorded cases per week per 100 currently admitted non-COVID-19 inpatients, during each phase of the study based on case report forms.

- Identification of linkage to individuals within an outbreak of SARS-CoV-2 nosocomial transmission using sequencing data that was not identified by pre-sequencing IPC evaluation, for each recorded HOCI during study phases in which the sequence reporting tool is in use.

4.2 Secondary outcomes

- Weekly incidence of IPC-defined SARS-CoV-2 hospital outbreaks, measured as incidence rate per week per 100 non-COVID-19 inpatients, during each phase of the study based on case report forms.

- Weekly incidence of IPC+sequencing-defined SARS-CoV-2 hospital outbreaks, measured as incidence rate per week per 100 non-COVID-19 inpatients, during each phase of the study based on case report forms.

- Occurrence or non-occurrence of changes to IPC actions implemented following receipt of SARS-CoV-2 sequence report, for each recorded HOCI during study phases in which the sequence reporting tool is in use.

- Record or no record of changes to IPC actions that would ideally have been implemented following receipt of SARS-CoV-2 sequence report, for each recorded HOCI during study phases in which the sequence reporting tool is in use.

- The number of HCW periods of sickness/self-isolation as assessed as a proportion of the number of staff usually on those wards impacted by HOCI cases, for all phases of the trial.

5 DATA

5.1 Data collection and management

Coded data will be collected from the study sites using paper and/or only electronic Case Record Forms (CRFs) and transcribed on to a CCTU database, stored on secure servers based at UCL. Data collection, data entry and queries raised by a member of the COG-UK HOCI study team will be conducted in line with the CCTU and study specific Data Management Standard Operating Procedure. Study documentation will be kept at the study site in a locked cabinet within a secured room. Clinical study team members will receive study protocol training. All data will be handled in accordance with the Data Protection Act 2018.

Data will be entered in the approved COG-UK HOCI database by a member of the clinical study team at site and protected using established CCTU procedures. Participants will be allocated a patient study identifier number (PIN). Data will be entered under this identification number onto the central database stored on the servers based at CCTU. The database will be password protected and only accessible to members of the COG-UK HOCI study team at CCTU, and external regulators if requested. The servers are protected by firewalls and are patched and maintained according to best practice. The physical location of the servers is protected by CCTV and security door access.

The database and coding frames have been developed by members of the COG-UK HOCI Trial Management Group and the Expert Data Sequencing Group, in conjunction with CCTU. The database software provides a number of features to help maintain data quality, including: maintaining an

audit trail, allowing custom validations on all data, allowing users to raise data query requests, and search/export facilities to identify validation failure/ missing data.
After completion of the study the database will be retained on the servers of UCL for on-going analysis of secondary outcomes, subject to the COPI notification by the UK Government remaining in force. Following expiration of the COPI notice, either Confidentiality Advisory Group (CAG) approval will be sought to continue to hold unconsented patient data, or the dataset will be anonymised and transferred into a STATA and/or comma separated value (CSV) format.

The enrolment logs, linking participant identifiable data to the pseudoanonymised PIN, will be held locally by the study site. This will either be held in written form in a locked filing cabinet or electronically in password protected form on hospital computers. After completion of the study the enrolment logs will be stored securely by the sites for 10 years unless otherwise advised by CCTU.

5.2 Data completion and schedule

Data entry for records of new HOCI cases will end by the end of May 2021.

5.3 Analysis dataset

Database lock will only occur once the last participant's last CRFs have been returned, data cleaning has been completed and all data queries are closed. A frozen dataset will be created by the study Data Manager for statistical analysis.

5.4 Data verification

Basic data checks are performed by the trial Data Manager periodically during the trial. Additional range, consistency and missing data checks will be performed when the datasets for analysis are constructed, as appropriate, before the statistical analysis is performed. All variables will be examined for unusual, outlying, unlabelled or inconsistent values.

Any problems with study data will be queried with the Trial Manager or Data Manager as appropriate. If possible, data queries will be resolved; although it is accepted that due to administrative reasons and data availability a small number of problems will continue to exist. These will be minimised.

5.5 Data coding

Details of the variables, including variable coding lists are included in the metadata which forms part of the trial master file.

5.6 Coding of primary and secondary outcomes

5.6.1 Primary outcome 1

Incidence of IPC-defined SARS-CoV-2 HAIs

In order to standardise this measure across sites, 'IPC-defined SARS-CoV-2 HAIs' will be considered to be all HOCIs with an interval of ≥8 days from admission to symptom onset (if known) or sample

date (i.e. those meeting the PHE definition of a probable or definite HCAI[1]). Incidence will be expressed 'per 100 non-COVID-19 inpatients per site per week', and will be evaluated for study baseline and intervention phases.

5.6.2 Primary outcome 2

Identification of SARS-CoV-2 nosocomial transmission using sequencing data

For each HOCI case during the intervention phases, the occurrence of this outcome will be defined as positive where the following two answers have been recorded in the Hospital Transmission section of CRF04:

> "Is sequencing report suggestive that patient is part of a hospital outbreak (i.e. involving ≥2 patients or HCWs in the hospital)?: Yes"
> &
> "If yes, was linkage to one or more of these patients suspected at initial IPC investigation?: No"

The occurrence of this outcome will be considered to be negative if the following answer is recorded:

> "Is sequencing report suggestive that patient is part of a hospital outbreak (i.e. involving ≥2 patients or HCWs in the hospital)?: No"

Or if the following combination is recorded:

> "Is sequencing report suggestive that patient is part of a hospital outbreak (i.e. involving ≥2 patients or HCWs in the hospital)?: Yes"
> &
> "If yes, was linkage to one or more of these patients suspected at initial IPC investigation?: Yes"

The outcome will be considered missing if either the first question is not answered, or if the first question is answered 'Yes' and the second question is not answered or is answered 'unknown'.

The outcome will also be considered negative if the viral sequence and sequence report have not been returned during the period of study data collection.

This outcome will only be evaluated for study sequencing intervention periods.

5.6.3 Secondary outcome 1

Incidence of IPC-defined SARS-CoV-2 hospital outbreaks

An IPC-defined SARS-CoV-2 hospital outbreak will be defined to occur when there are at least two HOCI cases on the same ward, with at least one having an interval of ≥8 days from admission to symptom onset (if known) or sample date, and with the outbreak event considered to be concluded

if there is a gap of 28 days before the observation of another HOCI case[1]. This will be evaluated using the ward location recorded at patient registration into the study (CRF01) for HOCI cases. The outbreak event will be considered to have occurred on the date of diagnosis of the first HOCI case. This outcome will be evaluated for study baseline and intervention phases.

5.6.4 Secondary outcome 2

Incidence of IPC+sequencing-defined SARS-CoV-2 hospital outbreaks

An IPC+sequencing-defined SARS-CoV-2 hospital outbreak will be defined to occur when there are at least two HOCI cases on the same ward that form a genetic cluster with maximum viral sequence pairwise SNP distance of 2 between each individual included and their nearest neighbour within the cluster. This will be evaluated using the ward location recorded at patient registration into the study (CRF01), with HOCI cases sorted into outbreak groups using the lists of close sequence matches on unit-ward as returned by the SRT and recorded in CRF03.

The outbreak event will be considered to have occurred on the date of diagnosis of the first HOCI case. This outcome will be evaluated for study sequencing intervention periods for all sites, and will also be evaluated for the baseline phase for sites with sequences available for HOCI cases in this phase.

5.6.5 Secondary outcome 3

Changes to IPC actions following receipt of sequencing report

For each HOCI case, the occurrence of this outcome will be defined as positive if 'Yes' is the answer to either of the following two questions in the 'Sequencing report impact on IPC team' section of CRF04:

> "Overall, did the sequencing report change IPC practice for this ward?"
> And/or
> "Has the sequencing report information been used in IPC decisions beyond this patient's ward?"

The occurrence of the outcome will be considered negative if at least one of these questions is answered 'No' and neither is answered 'Yes', and it will be considered missing if neither are answered.

A descriptive summary of relevant questions on CRF04 will be provided, without formal statistical analysis, among those patients for whom this outcome is recorded as positive.

The outcome will also be considered negative if the viral sequence and sequence report have not been returned during the period of study data collection.

This outcome will only be evaluated for study sequencing intervention periods.

5.6.6 Secondary outcome 4

Ideal changes to IPC actions following receipt of sequencing report

A binary outcome will be defined for each HOCI patient. This will be based on the value of Secondary Outcome 3, but will be additionally be defined as positive (whether Secondary Outcome 3 is negative or missing) if an 'increase' or 'decrease' that was not implemented is recorded for any of the actions in the 'other recommended changes to IPC protocols' section of CRF04.

A descriptive summary will be provided of all recorded recommended changes to IPC practice that were not implemented.

This outcome will only be evaluated for study sequencing intervention periods.

5.6.7 Secondary outcome 5

HCW sickness

The proportion of HCWs on sick leave due to COVID-19 will be calculated using the 'Current staffing levels on ward' section of CRF02. Analysis will be performed using the first available data within each IPC-defined SARS-CoV-2 hospital outbreak (as per secondary outcome 1), so as to provide a measure of the level of staff absence at the start of each outbreak. This outcome will be evaluated for study baseline and intervention phases.

5.7 Coding of adjustment and exposure variables

5.7.1 Total inpatients and proportion with COVID-19

This will be evaluated on a 'per study site per week' basis. Values from CRFs for each site will be compared for consistency within each week of data, with any outliers (reflecting data entry errors) excluded and the means taken of remaining values. Data on the CRFs for individual HOCIs may need to be supplemented with information from regular surveys sent to the trial teams or retrospective data extraction at each site to ensure completeness and accuracy.

5.7.2 Proportion of HCWs vaccinated for SARS-CoV-2

This will be evaluated on a 'per study site per fortnight' basis. We plan to obtain data through 2-weekly surveys of the trial team at each site.

5.8 Coding of calendar time

Analysis will be conducted with calendar time divided into 'study weeks', running from Monday-Sunday for all sites.

6 SAMPLE SIZE ESTIMATION

The following sample size calculations were included in the original protocol, written when the study was planned to commence in Summer 2020:

> The planned sample size is 4 sites all of which will implement rapid testing, followed by slower turnaround testing, followed by a second period without sequence data.
>
> There is uncertainty in the number of HOCIs that will be identified at each site during each of the intervention periods, with the rapid testing phase being roughly 8 weeks' duration. We assume there may be an average of 40 HOCIs/week per site, a total of 320 per site. Within a typical site this will allow us to estimate the proportion of HOCIs with genotypic linkage to another case(s) not detected by IPC processes with minimum precision of +/- 5.5%. Similarly we can estimate the proportion of HOCIs where an action is taken that would not have occurred without sequencing within +/-5.5%. We shall also calculate a pooled estimate of key proportions across the 4 sites implementing rapid sequencing, leading to estimation within +/- 11.3% for previously undetected linked cases for HOCIs and actions taken that would not have occurred, assuming an intracluster correlation coefficient of 0.05.
>
> Within the internal pilot phase for rapid testing, we shall record outcomes from around 160 HOCIs assuming a 4 week intervention period at the first site to initiate. This will permit us to estimate the proportion of previously undetected linked cases for HOCIs within +/- 7.7%.
>
> Comparing the proportion of HOCIs with genotypic linkage to another case(s) not detected by IPC processes between rapid testing and delayed testing within each site, the study would have at least 80% power to detect a percentage point difference of 14% (two-sided test with alpha=0.05, considering proportions of 57% vs 43% which would be associated with minimum power for a difference of this magnitude).
>
> If a second control phase is conducted of 4 weeks and assuming in this phase 50 IPC-defined HOCIs per week are observed in a site then, using an approximate Normal distribution for weekly counts, there is 80% power to demonstrate a reduction due to intervention of 13 IPC-defined HOCIs per week, under 5% significance level two-tailed testing. Pooling data across 4 sites and assuming the same reduction in all, there is 80% power to demonstrate a reduction of 6 IPC-defined HOCIs per week. In practice power will however be lower in the pooled analysis due to heterogeneity between sites. The same power calculations would also apply to the incidence rates of IPC-defined hospital outbreaks, although the study will have less power to detect a difference for this outcome due to the lower number of distinct outbreak events.

Substantial revisions to the required sample size were required with the study start date shifted to October 2020, with substantial variations in both community transmission and nosocomial transmission of SARS-CoV-2 across the UK at this point in time, with overall incidence of HOCIs lower than during the first wave of the pandemic in March-May 2020. The number of participating sites has been increased from 4 to 14 sites.

There is uncertainty in the number of HOCIs that will be identified at each site during each of the intervention periods, with the rapid sequencing phase being 8 weeks' duration. We assume there may be an average of 10 HOCIs/week per site during this intervention period, a total of 80 per site. Within a typical site this will allow us to estimate the proportion of HOCIs with genotypic linkage to another case(s) not detected by IPC processes with minimum precision of +/- 9.4%. Similarly we can estimate the proportion of HOCIs where an action is taken that would not have occurred without sequencing within +/-9.4%. We shall also calculate a pooled estimate of key proportions across the 14 sites implementing rapid sequencing, leading to estimation within +/- 6.5% for previously undetected linked cases for HOCIs and actions taken that would not have occurred, assuming an intracluster correlation coefficient of 0.05.

Comparing the proportion of HOCIs with genotypic linkage to another case(s) not detected by IPC processes between rapid testing and delayed testing phases across all sites, the study would have at least 80% power to detect a percentage point difference of 11% (two-sided test with alpha=0.05, considering proportions of 55.5% vs 44.5% which would be associated with minimum power for a difference of this magnitude).

For the outcome of weekly incidence of IPC-defined HOCIs, using an approximate Normal distribution for weekly counts there is 86.7% power to demonstrate a reduction from 12 IPC-defined HOCIs per week in the baseline phase to 10 per week during the rapid testing phase across all sites, under 5% significance level two-tailed testing. However, these calculations correspond to a variance of 12 for weekly counts based on the Poisson distribution, but the presence of over-dispersion of weekly counts would lead to a lower power to detect a difference. Using an overdispersion parameter of 0.82 based on retrospective analysis of data from Sheffield and Glasgow (dataset as described by Stirrup et al.[2]) results in 81% power to detect a reduction in mean weekly incidence from 12.5 to 10.

7 ANALYSIS PRINCIPLES

7.1 Intention-to-treat (ITT) or per-protocol?

The main analysis for the primary outcome of weekly incidence of IPC-defined SARS-CoV-2 HAIs and secondary outcomes of IPC-defined and IPC+sequencing-defined SARS-CoV-2 hospital outbreaks will be carried out on an 'intention-to-treat' (ITT) basis according to the defined study phases (i.e. regardless of the site-level performance in implementing the intervention as planned).

The secondary outcomes relating to the impact of sequencing on identification of nosocomial infection and on IPC actions will also be first evaluated on an ITT basis for all HOCI cases recorded within the study intervention periods. However, analyses will also be presented for these outcomes among only those HOCI cases with a sequence summary report returned within the target timeframe for each study phase.

7.2 Confidence Intervals and P-Values

All confidence intervals will be 95% and two-sided. Statistical tests will use a two-sided *p* value cut-off of 0.05 for statistical significance, with likelihood ratio tests (LRTs) used to calculate *p* values for treatment effects. Quadrature points for maximum likelihood estimation of generalised linear mixed effects models will be incremented until stability of parameter estimates and log-likelihood are achieved to ensure the validity of LRTs. The significance level of secondary outcomes will not be adjusted for multiple testing.

7.3 Baseline comparability

'Baseline' characteristics (i.e. patient characteristics at study entry) of recorded HOCI cases will be summarised by study intervention phase. These would not be expected to necessarily be consistent across phases, given that the intervention is at the site-level and the lack of randomisation of study phases.

7.4 Adjustment for design and contextual factors

We will account for the structure of the data with hierarchical random effects for each study site and each study phase within each study site.

Incidence and HCW absence outcomes will be adjusted for the proportion of inpatients at that point in time that were COVID-19 cases at each study site, included in regression models as covariates, whilst the number of non-COVID inpatients will be considered as an exposure variable (e.g. contributing to 'person-time' at risk of nosocomial infection). The cumulative proportion of HCWs vaccinated at each site for each study week and calendar time will also be included as covariates.

The analysis of per-HOCI outcomes will not use any adjustment variables.

7.5 Losses to follow-up and missing data

The primary outcome of weekly incidence of IPC-defined SARS-CoV-2 HAIs and secondary outcomes of IPC-defined and IPC+sequencing-defined SARS-CoV-2 hospital outbreaks rely on consistent recording of all eligible HOCI cases at each study site, but require only relatively minimal data completion for each HOCI case for these analyses (i.e. ward, admission date and COG-UK ID).

For the 'per HOCI' primary and secondary outcomes, data will be analysed for those cases with data fields completed as described in the 'Coding of primary and secondary outcomes' subsection, with levels of missingness reported.

The trial team will make every effort to obtain the data necessary for modelling of the incidence outcomes, which will be included in the analyses on a per week per site basis. If data are missing for any isolated weeks at any given site, then we will consider interpolation to impute vales based on the available data for adjacent weeks.

7.6 Summarising models

For the incidence rate outcomes, intervention effect estimates will be presented as incidence rate ratios and 95% CIs (generated using the Wald method) comparing each intervention phase to baseline.

For the binary outcomes defined at the 'per HOCI' level, the estimated marginal proportion obtained from the mixed effects logistic regression model will be reported for each phase with 95% CI, and the difference between the proportions in the slow and rapid intervention phases will be calculated with 95% CIs.

For analysis of the proportion of HCWs absent due to COVID-19, the estimated marginal proportion obtained from the mixed effects logistic regression model will be reported with 95% CI, and the difference between the proportions in the slow and rapid intervention phases versus the baseline phase will be calculated with 95% CIs.

The topic of small sample corrections for cluster randomised and other cluster-structured studies (e.g. stepped wedge trials) with outcomes that are not normally distributed is an area of ongoing active research. To our knowledge, there do not exist any studies regarding appropriate corrections for clustered data when analysing an outcome with negative binomial distribution. However, when calculating P-values and confidence intervals for the primary and secondary outcomes we will use a *t*-distribution with 12 or 13 degrees of freedom (*n* clusters – *n* relevant parameters) in order to ensure that there is not an inflated type-1 error rate. This correction has shown appropriate characteristics in simulation studies of analyses of binary outcomes using mixed effects models and generalised estimating equations[3, 4].

7.7 Descriptive summaries

Whilst formal statistical analysis of the per HOCI primary and secondary outcomes will only be conducted as specified, descriptive summaries of a more complete set of data items from the study CRFs will also be provided (planned dummy tables are included in the SAP Section 9).

8 ANALYSIS DETAILS

8.1 Recruitment, intervention uptake and follow-up

Details will be provided of the duration and order of the baseline and intervention phases at each study site, and of the numbers of HOCI cases recorded during each study period-phase combination.

A summary will also be produced of the proportion of HOCI cases for whom a viral sequence and/or sequence summary report is returned to IPC teams within the target timeframe for each study phase, alongside the median, IQR and range of the time to return of the report.

8.2 Baseline Characteristics

Baseline characteristics of recorded HOCI cases will be summarised by study intervention phase. Summary measures for the baseline characteristics of each phase will be presented as median and interquartile range for age, and frequencies and percentages for categorical variables.

8.3 Analysis Methods

Analyses will be conducted using Stata V16 unless specified otherwise.

8.3.1 Random effects structures

All random effects models will include a random intercept term per study site, and exchangeable random effect terms for each study phase within each site. All random effects will be specified as normally distributed.

8.3.2 Incidence outcomes

Incidence outcomes for the occurrence of IPC-defined SARS-CoV-2 HAIs and secondary outcomes of IPC-defined and IPC+sequencing-defined SARS-CoV-2 hospital outbreaks will be analysed on a weekly basis per 100 non-COVID-19 inpatients (i.e. with an exposure variable representing total non-COVID inpatient-weeks/100 for each data entry). Mixed effects negative binomial models will be used for these outcomes.

Data for incidence outcomes in the first week of each intervention period, or in the first week of return to an invention following a break (e.g. over the Christmas and New Year period) will be assigned an indicator variable defined according to the changeover (transition) period concerned, and so will not be considered as direct evidence regarding the intervention effect.

Exploratory graphical summaries of incidence outcomes over time for each site will also be produced.

8.3.3 Per HOCI outcomes

The primary outcome of identification of SARS-CoV-2 nosocomial transmission using sequencing data and the secondary outcomes relating to changes to IPC actions will be analysed using mixed effects logistic regression models. Marginal proportions from the fitted models will be reported for the rapid- and slow-turnaround intervention phases, and a descriptive evaluation of between-site variation will be provided including the median, IQR and range of observed values across sites. A mixed effects logistic regression model will also be used to evaluate the magnitude of any differences in these outcomes between the rapid- and slow-turnaround intervention phases.

8.3.4 HCW absence

The proportion of HCWs absent on sick leave due to COVID-19 will be analysed using a binomial form mixed effects logistic regression model. In addition to hierarchical random effect terms for each study site and each study phase within each study site, an additional random effect will be defined per ward.

8.4 Adjustment for contextual factors in analysis

For the weekly incidence and HCW absence outcomes, models will be adjusted by the proportion of current inpatients that are COVID-19 cases across the study site. This proportion will be included in the models as a continuous and potentially non-linear adjustment factor using a 5 knot restricted cubic spline[5]. The fitted relationship for these analyses will be reported in the primary study report.

We also plan to adjust the weekly incidence and HCW absence outcomes according to the level of vaccine roll-out to HCWs at each study site. If possible, this proportion will be included in the models as a continuous and potentially non-linear adjustment factor using a 5 knot restricted cubic spline. These outcomes may also be affected by vaccine roll-out across the general population. As vaccine deployment is planned to be consistent across the UK this will be accounted for by adjusting analyses for calendar time, again using a 5 knot restricted cubic spline in order to fit a smoothed relationship with time.

For adjustments related to both the COVID-19 inpatient proportion and vaccine roll-out, simpler relationship (e.g. linear adjustment) will be considered if the spline model does not converge or shows very large standard errors for estimated coefficients.

8.5 Sensitivity analyses

8.5.1 Exclusion of clinic-periods with suboptimal implementation

We will conduct sensitivity analyses excluding study sites and/or periods with suboptimal implementation of the trial intervention, both in terms of overall population sequencing coverage for HOCIs and the turnaround time for sequence reports being returned to IPC teams. The exact criteria for this will be decided amongst the study team before any analysis has been conducted.

8.5.2 Impact on IPC actions

For the Secondary Outcomes related to implemented and recommended changes to IPC actions, these will also be summarised and analysed using only the first HOCI case with data entered in each IPC+sequencing-defined SARS-CoV-2 hospital outbreak, reflecting the fact that a set of changes is likely to be made and then sustained following the first observation of infection on a ward.

If the target turnaround time for sequence generation and reporting is missed for a substantial proportion of HOCI cases in each of the intervention phases, then results will also be reported separately for the subset of cases for which the intervention was implemented within the target timeframe.

8.6 Regression diagnostics

We shall check for model stability primarily through examination of standard errors for covariates, with large standard error or convergence problems indicating a problem of model instability.

9 INTERIM ANALYSES

9.1 Regular reports to TSC-DMC

Reports to the TSC-DMC during the duration of the trial will be limited to overall case counts and basic demographic characteristics of included patients. This is due to the fact that the study does not involve an investigational medicinal product and, as such, the risk of a negative impact of the intervention on participants is low.

9.2 Decision regarding continuation of trial into final phase

A decision regarding the final phase of the study (Period 4) is planned for April 2021, with the options being: ending of the trial at Period 3, a further phase of rapid sequencing at each site or a further phase of 'baseline' date collection without use of the SRT. A recommendation regarding this decision will be made by the study investigators and agreed with the TSC-DMC. The decision will be determined by the course of the epidemic and the progress of vaccination among key risk groups, and by the quantity of data that has been collected and is expected to be collected by the end of Period 3 under each exposure in the study, and not based on interim evaluation of the effect of the sequencing intervention under investigation on the incidence of nosocomial infection. It is in any case likely that data entry and reporting delays for both patient-level data and hospital-level data (e.g. inpatient numbers and HCW vaccinations) would make a comprehensive assessment of intervention effect impossible within the required timeframe.

10 TABLES AND GRAPHS:

10.1 Tables

Table 1: Demographic and Baseline Characteristics of the recruited participants by study phase

	Trial phase			
Characteristic at screening	Baseline	Slow turnaround sequencing	Fast turnaround sequencing	**Total**
N HOCI cases				
HCAI classification, *n* (%)				
Indeterminate (3-7 days)				
Probable (8-14 days)				
Definite (>14 days)				
Age (years), median (IQR)				
Age ≥70 years, *n/N* (%)				
Sex at birth: female, *n/N* (%)				
Ethnicity, *n* (%)				
White British/Irish				
White - other				
Mixed ethnicity				
South Asian				
Chinese				
Other Asian				
Caribbean				
African				
Black - other				
Unknown				
Symptomatic at time of sampling, *n/N* (%)				
Significant comorbidity present, *n/N* (%)				
Pregnant, *n/N* (%)				
Hospital admission route, *n* (%)				
ED				
Hospital transfer				
Care home				
GP referral				
Outpatient clinic ref.				
Other				
Unknown				

Table 2: Per HOCI implementation and outcome summary by study intervention phase

	Trial phase			
	Baseline	Slow turnaround sequencing	Fast turnaround sequencing	**Total**
N HOCI cases				
Implementation				
Sequence returned within expected timeline, *n* (%)	—			
Sequence returned within study period, *n* (%)	—			
SRT report returned within expected timeline, *n* (%)	—			
SRT report returned within study period, *n* (%)	—			
Time from sample to report return (days), median (IQR, range)	—			
Sequencing results				
SRT suggestive patient acquired infection post-admission, *n/N* (%)	—			
SRT suggestive patient is part of ward outbreak, *n/N* (%)	—			
Linkage identified not suspected at initial IPC investigation, *n/N* (%*, 95% CI)	—			
SRT excluded IPC-identified hospital outbreak, *n/N* (%)	—			
Impact on IPC				
SRT changed IPC practice, *n/N* (%*, 95% CI)	—			
SRT changed IPC practice for ward, *n/N* (%)	—			
SRT used in IPC decisions beyond ward, *n/N* (%)	—			
IPC team reported SRT to be useful, *n/N* (%)				
Yes	—			
No	—			
Unsure	—			
SRT would ideally have changed IPC practice, *n/N* (%*, 95% CI)	—			
HCW absence on ward				
Prop. HCWs on sick leave due to COVID-19, median (IQR, *n*); mean* (95% CI)				

IQR, interquartile range; Prop., proportion. *Estimated marginal value from mixed effects model.

Table 3: Incidence outcomes by study intervention phase

	Trial phase			**IRR† (95% CI, *P*)**	
	Baseline	Slow turnaround sequencing‡	Fast turnaround sequencing‡	**Slow vs baseline**	**Fast vs baseline**
n HOCI cases				—	—
n IPC-defined HAIs				—	—
Weekly incidence of IPC-defined HAIs per 100 inpatients, mean (median, IQR, range)*					
n IPC-defined outbreak events				—	—
Weekly incidence of IPC-defined outbreak events per 100 inpatients, mean (median, IQR, range)*					
n IPC+sequencing-defined outbreak events	(dependent on available COG-UK IDs)			—	—
Weekly incidence of IPC+sequencing-defined outbreak events per 100 inpatients, mean (median, IQR, range)*	(dependent on available COG-UK IDs)			(dependent on available COG-UK IDs for baseline)	(dependent on available COG-UK IDs for baseline)

*Descriptive data over all week-long periods at all study sites. †Adjusted for proportion of current inpatients at site that are COVID-19 cases. ‡Not including data from the first week of each intervention period, or in the week following any break in the intervention period.

Table 4: Per outbreak event outcomes by study intervention phase

	Trial phase			
	Baseline	Slow turnaround sequencing	Fast turnaround sequencing	**Total**
IPC-defined outbreak events				
n outbreak events				
n/N (%) of HOCI cases part of outbreak event				
Number of HOCIs per outbreak event, median (IQR, range)				
Prop. HCWs on sick leave due to COVID-19, median (IQR, *n*); mean* (95% CI)				
IPC+sequencing-defined outbreak events				
n outbreak events	—			
n/N (%) of HOCI cases part of outbreak event	—			
Number of HOCIs per outbreak event, median (IQR, range)	(dependent on available COG-UK IDs)			
For first HOCI in outbreak:				
SRT changed IPC practice, *n/N* (%*, 95% CI)	—			
SRT changed IPC practice for ward, *n/N* (%)	—			
SRT used in IPC decisions beyond ward, *n/N* (%)	—			
IPC team reported SRT to be useful, *n/N* (%)				
Yes	—			
No	—			
Unsure	—			
SRT would ideally have changed IPC practice, *n/N* (%*, 95% CI)	—			

*Estimated marginal value from mixed effects model.

Table 5: Descriptive summary of impact of sequencing on IPC actions implemented during study intervention phases

	Trial phase					
	Slow turnaround sequencing			Fast turnaround sequencing		
N HOCI cases						
Review of IPC actions already taken	*Support*	*Refute*	*Missing*	*Support*	*Refute*	*Missing*
SRT results support or refute IPC actions already taken						
Changes to IPC practice following sequencing	*To enhanced*	*To routine*	*No change*	*To enhanced*	*To routine*	*No change*
Change to cleaning protocols on ward						
	To greater	*To fewer*	*No change*	*To greater*	*To fewer*	*No change*
Change to visitor restrictions						
	To 'cohort nursing'	*To 'no restrictions'*	*No change*	*To 'cohort nursing'*	*To 'no restrictions'*	*No change*
Change to staffing restrictions on ward						
	Increase	*Decrease*	*No change*	*Increase*	*Decrease*	*No change*
Hand hygiene audit frequency						
IPC staff visits to ward						
Assessment of alcogel stocks						
Assessment of soap stocks						
Assessment of aseptic non-touch technique compliance						
Assessment of PPE supply						
Availability of doffing and donning buddy						
IPC signage assessment						
IPC signage implementation						
Training on IPC procedures						

Data shown as *n* or *n/N* (%).

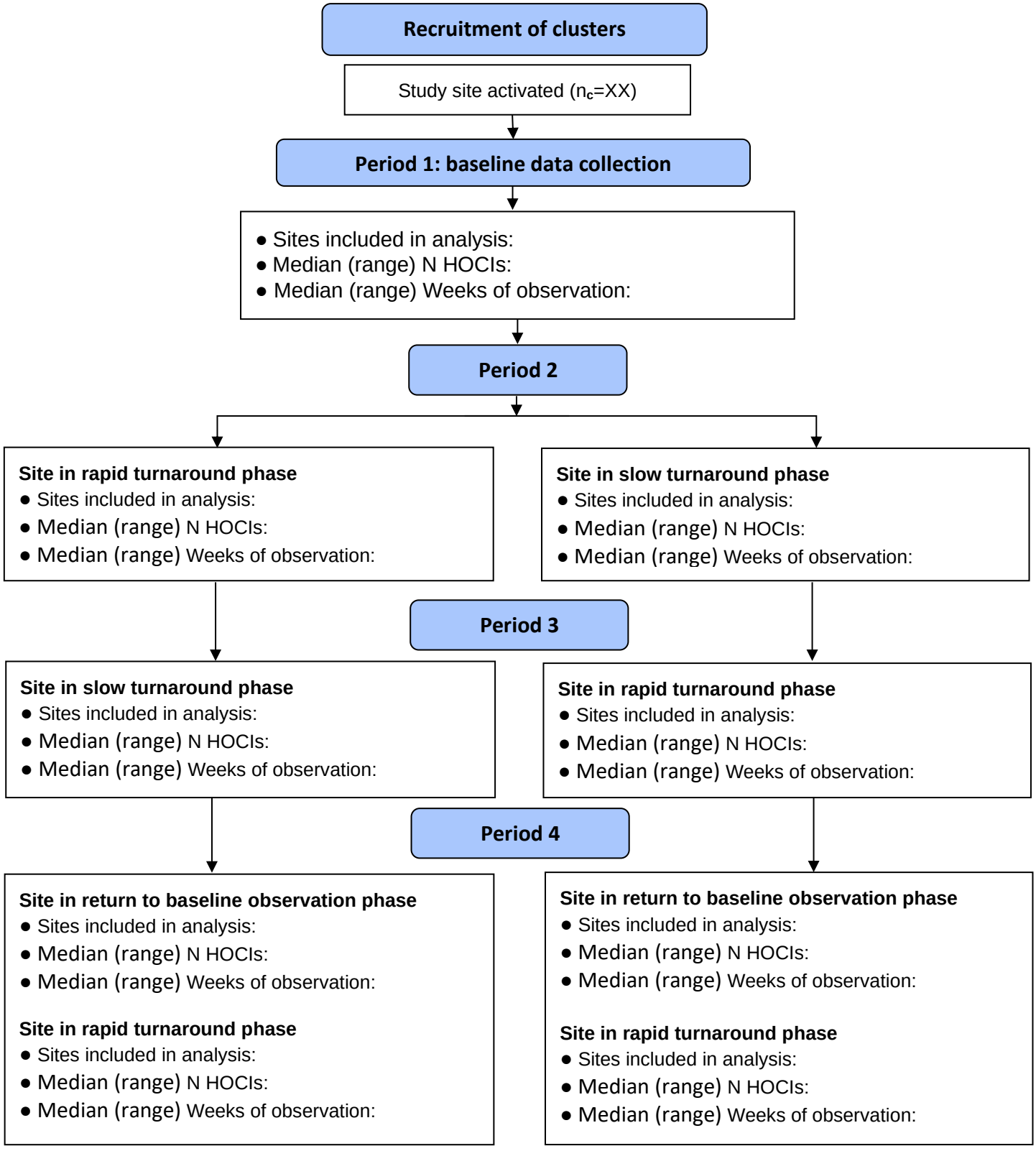
Table 6: Descriptive summary of impact of sequencing on IPC actions implemented during study intervention phases, only including the first HOCI in each IPC+sequencing-defined outbreak event

	Trial phase					
	Slow turnaround sequencing			Fast turnaround sequencing		
N HOCI cases						
Review of IPC actions already taken	*Support*	*Refute*	*Missing*	*Support*	*Refute*	*Missing*
SRT results support or refute IPC actions already taken						
Changes to IPC practice following sequencing	*To enhanced*	*To routine*	*No change*	*To enhanced*	*To routine*	*No change*
Change to cleaning protocols on ward						
	To greater	*To fewer*	*No change*	*To greater*	*To fewer*	*No change*
Change to visitor restrictions						
	To 'cohort nursing'	*To 'no restrictions'*	*No change*	*To 'cohort nursing'*	*To 'no restrictions'*	*No change*
Change to staffing restrictions on ward						
	Increase	*Decrease*	*No change*	*Increase*	*Decrease*	*No change*
Hand hygiene audit frequency						
IPC staff visits to ward						
Assessment of alcogel stocks						
Assessment of soap stocks						
Assessment of aseptic non-touch technique compliance						
Assessment of PPE supply						
Availability of doffing and donning buddy						
IPC signage assessment						
IPC signage implementation						
Training on IPC procedures						

Data shown as *n* or *n/N* (%).

10.2 Graphs

Graph 1: Flow diagram of study site enrolment and intervention implementation



11 REFERENCES

1. Public Health England. **COVID-19: epidemiological definitions of outbreaks and clusters in particular settings**. In. https://www.gov.uk/government/publications/covid-19-epidemiological-definitions-of-outbreaks-and-clusters; 2020.
2. Stirrup OT, Hughes J, Parker M, Partridge DG, Shepherd JG, Blackstone J, et al. **Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data**. *medRxiv* 2020:2020.2011.2012.20230326.
3. Li P, Redden DT. **Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials**. *BMC Medical Research Methodology* 2015; 15(1):38.
4. Thompson JA, Hemming K, Forbes A, Fielding K, Hayes R. **Comparison of small-sample standard-error corrections for generalised estimating equations in stepped wedge cluster randomised trials with a binary outcome: A simulation study**. *Statistical Methods in Medical Research* 2020:0962280220958735.
5. Kahan BC, Rushton H, Morris TP, Daniel RM. **A comparison of methods to adjust for continuous covariates in the analysis of randomised trials**. *BMC Medical Research Methodology* 2016; 16(1):42.