A phase 2 multi-centre adaptive randomised platform trial to assess antiviral pharmacodynamics in early symptomatic influenza infection (AD ASTRA): Statistical Analysis Plan

June 18, 2024

Registered at clinicaltrials.gov number **NCT05648448**

Version 1.0; refers to Master Protocol version V5.0

Written by:
Dr James Watson (Statistician)
Signature: Date: 19/08/2024

Reviewed by and approved by:

Prof Sir Nicholas White (Co-PI)
Signature: Date: 21/06/2024

Dr William Schilling (Co-PI)
Signature: Date: 22/06/24

Contents

1 Trial Overview

Before the COVID-19 pandemic, annual seasonal influenza epidemics (mainly associated with influenza A/H3N2, influenza A/H1N1 pandemic 2009 and influenza B strains), were estimated by the WHO to cause around 1 billion infections, 3-5 million cases of severe illness and 290,000-650,000 deaths each year. Several classes of antiviral drugs are approved in different countries for the treatment of influenza infection in those at risk of severe disease: neuraminidase inhibitors (five days of oral oseltamivir, single dose of intravenous peramivir, five days of inhaled zanamivir, single dose of inhaled laninamivir), an endonuclease inhibitor (single dose of oral baloxavir), a polymerase inhibitor (five days of oral favipiravir) and a haemagglutinin inhibitor (five days of oral umifenovir).

Despite the availability of a large number of antivirals to treat influenza, only one of these is recommended by the WHO and this recommendation is based on low-quality evidence for clinical outcomes. Although some clinical trials of new antivirals have included oseltamivir as a comparator arm, direct comparisons of antiviral and clinical efficacy between the multiple available antivirals are lacking. This comparative information is important for guideline development and aiding purchasing and prioritisation decisions when several options are available.

AD ASTRA is a multi-country platform adaptive trial which provides a methodology for the quantitative assessment of antiviral effects in low-risk patients with high viral burdens and uncomplicated COVID-19. We choose patients with a low risk of progression as this justifies a no treatment arm; we target patients with high viral loads as this is a subgroup in whom antiviral effects can be detected and quantified more easily.

This statistical analysis plan (SAP) covers analyses for both the interim and the final reports. We developed the SAP following the Guidelines for the Content of Statistical Analysis Plans in Clinical Trials [1]. It includes pre-specified decision rules for continuing or stopping individual trial arms based on effectiveness or futility. It is based on the SAP for the PLATCOV trial which has a near identical design, assessing the effectiveness of antiviral interventions for SARS-CoV-2 [2].

1.1 Main research questions

Each clinical site in this platform trial will simultaneously test multiple interventions (depending on local regulations and drug availability). There are initially two control arms. The negative control which is no study drug arm; this involves no intervention other than antipyretics; and the positive control which is baloxavir, chosen on the basis of the best in vitro and clinical results available [3]. Other interventions tested in the platform which are shown to have measurable in vivo antiviral effect will then be compared to the positive control as a non-inferiority comparison using a pre-specified threshold.

The primary objective of this trial is to characterise the antiviral activity of currently available drugs and those with potential activity by comparing their viral clearance dynamics with no treatment. Many of these drugs are already used and recommended in some countries. Showing that they do not have measurable antiviral activity is as important as showing that they do. For interventions that do have a clear measurable antiviral activity, we need to determine how good they are so policy makers can make informed choices.

The secondary objectives of this study are: 1) To characterise the determinants of viral clearance in early influenza infection e.g. contribution of baseline serology, influenza type/subtype, prior vaccination, host genetics; 2) To determine optimal dosing regimens for drugs shown to have considerable antiviral activity; 3) To compare time to symptom resolution and fever duration between interventions; 4) To determine the effects of drugs on the development of drug resistant viral mutants between interventions and no treatment arm.

The tertiary objectives of this study are: 1) To characterise the relationship between viral clearance and hospitalisation for clinical reasons by day 28; 2) To characterise the relationship between viral clearance and development of influenza complications including bronchitis, sinusitis, otitis media and pneumonia requiring antibiotics by day 28; 3) To Characterise the relationship between viral clearance, randomisation arm and other measures (covariates) and development of post-acute symptoms in comparison to post-acute symptoms for other respiratory viral diseases (e.g. long-COVID).

1.2 Outcome definitions

Primary outcome

- Rate of viral clearance estimated from the $\log_{10}$ viral density derived from qPCR of standardised duplicate oropharyngeal swabs taken daily from baseline (day 0) to day 5.

Secondary outcomes

- Time to resolution of fever, defined for the patients with a fever at baseline (at least one axillary temperature measurement within the first 24 hours from randomisation $\geq$ 37.5). Resolution of fever is defined as an axillary temperature $\leq$37.0C for both measurements taken over one day (at least 24 hours).
- Time to resolution of symptoms, defined as first day with no reported symptoms.

Tertiary outcomes

- Hospitalisation for clinical reasons up to day 28.
- Development of influenza-related complications including bronchitis, sinusitis, otitis media and pneumonia requiring antibiotics, up to day 28.

2 Study Methods

2.1 Trial design

This is a multi-centre, multi-country, open-label, randomised, controlled, adaptive platform trial of antiviral interventions in early symptomatic influenza. Interventions currently included in the platform are: baloxavir, favipiravir, and oseltamivir. Baloxavir has been designated as the positive control. Randomisation is uniform across the interventions.

2.2 Randomisation

Randomisation is performed via a centralised web-app designed by MORU research software engineers using RShiny, hosted on a MORU server. Each study nurse responsible for randomising patients has unique login credentials provided by email from the main study statisticians (James Watson or Phrutsamon Wongnak). Randomisation sheets are pre-generated for each study site separately using blocks of size K (where K is equal to 3 times the number of interventions available at the site) with an additional 100/K% fuzziness (randomly interchanging one of the allocations per block to avoid the study nurse knowing exactly what the Kth allocation will be).

Randomisation sheets and randomisation event logs are stored on a secure Dropbox folder (professional version that has full version control) which is accessed directly by the RShiny app. Only the study statisticians James Watson and Phrutsamon Wongnak and the MORU IT manager have read/write access privileges to this Dropbox folder. For cross checking purposes, the randomisation app also records the patient age and sex. Each randomisation event is logged with the corresponding time-stamp and the identity of the nurse who performed the randomisation. Each time the set of interventions available at a given study site changes, a new randomisation sheet is generated, overwriting the previous one.

2.3 Sample size projections

The sample size is adaptive (there is no fixed sample size). For each intervention the final sample size is based upon pre-specified stopping rules which use 1) a futility/success margin λ_1 for comparisons with the no study drug arm, and 2) a non-inferiority margin λ_2 for comparisons with the positive control arm. Because there are theoretical boundary cases whereby the sample size would need to be very large in order to meet the stopping rule (when the effect is exactly equal or very close to the margin), there is a maximum sample size of 120 patients per intervention arm. Each intervention arm will be stopped before N = 120 if it meets the following stopping rules in the interim analyses, unless in discussion with the Data Safety and Management Board (DSMB) or Trial Steering Committee (TSC) there is a good rationale and decision to continue recruitment further:

1. Meets the futility criteria: if Probability[effect $< \lambda_1$] > 0.9;

2. 1) Meets the success criteria: if Probability[effect $> \lambda_1$] > 0.9 relative to the no-study drug, and 2) meets the non-inferiority criteria or the inferiority criteria: if Probability[effect $> -\lambda_2$] > 0.9 or Probability[effect $< -\lambda_2$] > 0.9 relative to the positive control arm, respectively.

Stopping for success requires both meeting the success criteria relative to the no study drug arm and concluding either non-inferiority or inferiority relative to the positive control arm, with comparisons done in that order.

The average sample sizes thus depend on the values of λ_1 and λ_2. Given the model-dependent nature of the analysis, simulation is required to estimate the expected distribution of sample sizes for varying effect sizes. The following sections outline how values for λ_1 (and now λ_2) were selected and changed over the course of the study.

2.3.1 Sample size simulations

The determinations of the futility/success margin λ_1 and the non-inferiority margin λ_2 were informed by insights gained from a series of simulations using data from the first 500 patients enrolled in the PLATCOV trial, which has a similar study design for assessing the efficacies of SARS-CoV-2 antivirals.

The simulations assessed thresholds $\lambda_1 \in \{1.125, 1.15, 1.2\}$ (i.e. 12.5%, 15% or 20%) for the futility/success assessments relative to the no study drug arm; and thresholds $\lambda_2 \in \{0.905, 0.875, 0.85\}$ (i.e. -9.5%, -12.5%, or -15%) for the non-inferiority assessments relative to the positive control arm. We used the linear model fit to the first 500 patients to simulate trial data. 100 iterations were done for each permutation of λ_1, λ_2 and for three hypothesised effect sizes (0, 40% and 60% increase relative to no study drug), i.e. 2700 simulations in total. Preliminary simulations showed that most false positive or false negative results (stopping early for futility an intervention that has effect of 40 or 60%, or stopping early for success an intervention that has an effect of 0%) occurred for interim analyses done on 10 patients per arm. For this reason, we updated the interim analysis plan so that the first interim analysis is triggered at 20 patients per arm, with subsequent analyses for each additional 10 per arm. For the non-inferiority comparisons, they start when there are at least 40 patients per arm, and only if the success stopping rule has been met (Figure 1). Each model fit only used simulated data from the intervention and the relevant control arm (the negative control for the futility/success stopping rules, the positive control for the non-inferiority stopping rules).

Choosing optimal λ_1 The results of the simulations for the success and futility stopping rules are shown in Supplementary Figures S1-S2. Comparable results were obtained as for the previous sample size simulations [4]. For $\lambda_1 = 1.125$ (12.5%) and an intervention that had no effect, the median sample size was 40 patients per arm with 4% false positive results and 28% of inconclusive trials at 120 patients per arm. Increasing λ_1 to 1.2 (20%) would reduce the median sample size to 30 patients per arm and result in approximately 1% false positive results and 18% of inconclusive trials at 120 patients per arm.

For an intervention with an effect of 40%, the median sample size was 30 patients per arm for a threshold of 12.5% and 50 patients per arm for a threshold of 20%. In all cases, false negative results were less than 3% of simulations, and inconclusive results increased from 14% for a threshold of 12.5% to 34% for a threshold of 20%.

For an intervention with an effect of 60%, the median sample size was 20 patients per arm, regardless of the λ_1 threshold. False negative results were less than 1% of simulations, and inconclusive results increased from 1% for a threshold of 12.5% to 9% for a threshold of 20%.

Following discussion with the trial steering committee for the PLATCOV trial, it was decided to increase the λ_1 threshold to 20% in order to stop poorly performing arms earlier (this reflects how the therapeutic priorities of the trial have changed as the pandemic has progressed).

Choosing optimal λ_2 The results of the simulations for the non-inferiority stopping rules are shown in Supplementary Figures S3-S4.

In simulations which met the stopping rule for success (Probability[effect$>\lambda_1$]>0.9), stopping for non-inferiority or inferiority was then assessed for each of the three threshold values. Only in 1 simulation with an effect of 0% was the non-inferiority threshold met ($<$0.3% false positive result).

An effect of 40% relative to no study drug arm is very close to/on the boundaries defined by the λ_2 values, assuming that the positive control has an effect of 60% relative to the no study drug.

For all three boundary values, around half the simulations met the non-inferiority criterion, 10% were inconclusive and 40% met the inferiority criterion.

For an effect of 60% relative to the no study drug arm (i.e. same as the positive control), 70% to 78% of simulations met the non-inferiority criterion for the -9.5% (i.e. most stringent) to the -15% (least stringent) thresholds, respectively. Around 6% of simulations were inconclusive at 120 patients, and 24 to 17% resulted in a false inferiority result for the -9.5% (i.e. most stringent) to the -15% (least stringent) thresholds, respectively.

Following discussion with the trial steering committee for the PLATCOV trial, it was decided to initially set the λ_2 threshold at -10%. This results in earlier stopping of interventions which are clearly less effective than the positive control and later stopping for interventions with an effect close to the positive control.

2.4 Framework

All primary analyses and stopping decisions will be made using models including data from **contemporaneous controls only**. As this is an adaptive trial, temporal drift (due to mutations in the virus, changes in the patient characteristics, etc.) can confound comparisons with non-contemporaneous controls [5]. However, we will also fit models to the full data set (using all controls), with flexible spline fits to adjust for temporal trends.

All primary comparisons for futility/success will be made with respect to the negative control arm; all primary comparisons for non-inferiority will be made with respect to the positive control arm (no "across intervention" arm comparisons in the primary analysis). In the case of a site not randomising patients to the negative control or the positive control (because of local objection or availability), that site will only provide indirect data to support the treatment estimates (under the hierarchical model structure). All decisions concerning efficacy will be based on super-superiority relative to the negative control arm (λ_1) whereby the probability of an effect is defined under the model as the posterior probability that the increase in viral clearance relative to the control arm is greater than λ_1. λ_1 is initially chosen as 20%.

For interventions which show a measurable antiviral effect (i.e. meet the stopping rule for success), we will then compare their antiviral effect to the positive control arm (baloxavir), using the non-inferiority threshold of λ_2. λ_2 is initially chosen as 10%.

3 Statistical interim analyses and stopping rules

We plan frequent interim analyses (Figure 1). This is to allow for near-real time monitoring of accrued data in order to detect possible issues in patient recruitment, viral load swabbing, or with the PCR assays at the different sites. The stringent margin for success (greater than 20% increase) results in a very low type 1 error. In addition, meeting the success threshold triggers a subsequent non-inferiority comparison with the positive control arm and where deemed appropriate an intensive PK-PD sub-study to determine dose-response effects in order to gather additional pharmacometric data to inform dosing for that intervention.

3.1 Frequency and timing of the interim analyses

Decision making analyses will be done by arm (only using the data from that arm and the concurrent negative or positive controls). The first interim analysis for a given arm will be performed when a minimum of 20 patients randomised to that arm (and 20 concurrent negative controls) have available and analysable qPCR data (PCR data will lag behind actual recruitment, usually by approximately 3 weeks). Subsequent analyses will be performed at most for every additional 10 patients (10 randomised to the intervention and 10 to the control arm).

3.2 Interim analysis decision making

Figure 1 shows the decision algorithm for removing an intervention from the platform or replacing the positive control. Only the study statisticians (James Watson and Phrutsamon Wongnak) will be unblinded to all interim analyses. Each interim analysis report will be sent to the DSMB along with a summary of whether the futility or success thresholds for any of the interventions have been met. The DSMB will then make recommendations for continuing or stopping the intervention

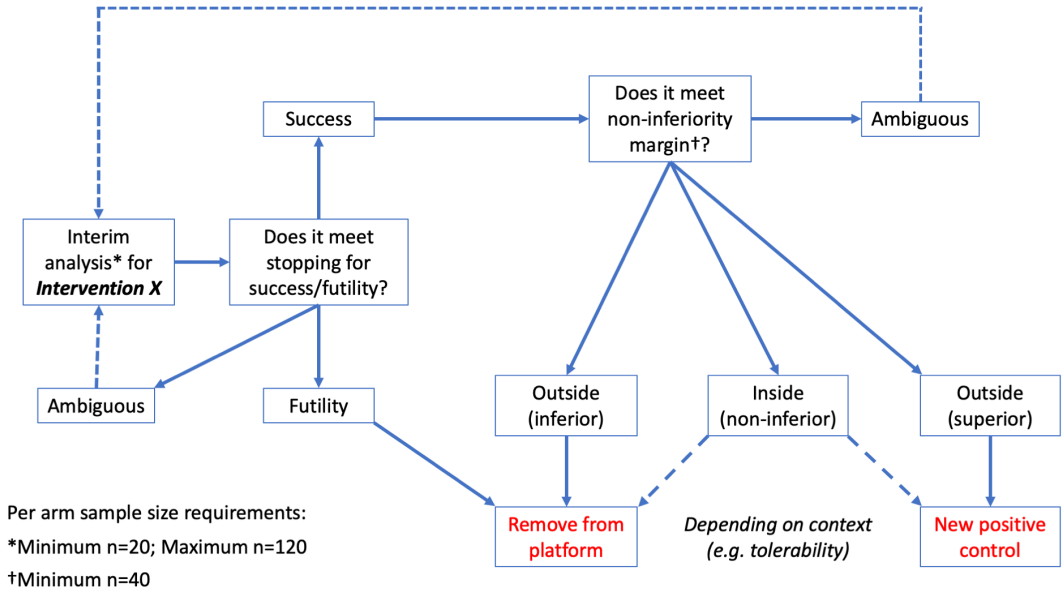

Figure 1: Planned interim analyses and decision rules for removing an arm from the platform or making an arm the new positive control. The success threshold λ_1 is defined as 20% increase in the rate of viral clearance.

arms. Concurrent data from other trials (e.g. safety data) may be used in making these decisions. If an arm is stopped for futility or an intensive PK-PD study is triggered, then all the data pertaining to the negative control arm and the intervention arm will be presented to the Trial Steering Committee (TSC), along with the DSMB recommendations. An unblinded analysis of the first 50 patients with available qPCR data will be presented to the TSC in order to determine any necessary changes to the statistical analysis plan.

3.3 Timing of final analysis

As this is a platform trial, no overall final analysis is planned. For each intervention studied, we define the "final analysis" as the point when the last patient randomised to that arm passes their day 28 follow-up.

3.4 Timing of outcome assessments

For the primary and secondary endpoints we will use all viral load measurements taken up until day 5. Time will be defined as time since randomisation (units of days, including all timepoints $<$ 5.5 days since randomisation as there will be some variation in exact clock times of the follow-up swabs).

For fever and symptom resolution, time to resolution will be calculated from daily measurements taken over the first week with right censoring for unobserved events (withdrawal or no resolution by 7 days).

For the tertiary endpoints, all cause hospitalisation and development of complications will be taken up until day 28.

4 Statistical Principles

4.1 Posterior estimates

Treatment effects will be estimated under a Bayesian framework using a hierarchical model with weakly informative prior distributions. For the hyperparameters, the prior distributions are chosen for computational reasons to aid model convergence. For the key population level parameters, we will use priors based on kinetics of viral clearance of SARS-CoV-2. At the first interim analysis

these can be changed if the observed kinetics diverge substantially. The prior on the treatment effect is driven by plausible maximum effect sizes given known effect sizes in SARS-CoV-2 from the PLATCOV trial.

4.2 Adherence and protocol deviations

This study is open label and thus it is possible that some patients who are randomised to the no study drug arm, or to study interventions perceived to be ineffective, may be given alternative rescue treatment if their symptoms persist and the treating physician is worried that they will deteriorate clinically. This can introduce confounding between the treatment allocation and outcome. Notably, if multiple patients randomised to the negative control were in fact given the positive control before day 5, this could impact the viral loads during follow-up and thus bias the estimates of the rate of viral clearance in the negative control arm. For all protocol deviations regarding treatment (either stopping treatment for safety reasons or switching treatment arms), we will consider all viral load measurements taken after the deviation as censored, i.e. we will estimate viral load clearance rates only using the data during the period of protocol adherence. We will summarise the number of treatment protocol deviations by site.

4.3 Analysis populations

The primary analysis will be in a modified intention-to-treat (mITT) population defined as all patients who have a baseline viral density greater than 250 genome copies per ml. This excludes from the primary analysis patients with very low viral densities who could have different kinetics. If a patient is given rescue treatment, or changes treatment during follow-up (in the first 5 days), then the primary analysis will only include follow-up data until day of switch.

A second sensitivity analysis will be conducted using data from all randomised patients (including those who have all negative samples; under the Bayesian model the slope will be shrunk towards the population mean value).

The safety population will include all patients who have received at least one dose of the intervention.

5 Trial population

Figure 2 shows an example CONSORT diagram summarising the number of patients screened; the reasons for exclusion; the number of enrolled patients randomised to each arm; and the number of patients who complete treatment and follow-up and who are included in the modified intention to treat analysis (mITT).

5.1 Screening data

We will summarise the reasons for exclusion for the screened patients not eligible for enrolment. This will not be done for the interim analysis, only for the final analysis once an arm is stopped for futility or success.

5.2 Eligibility

The key trial eligibility criteria are as follows:

- Previously healthy adults, aged 18 to 60 years (low risk of developing severe influenza symptoms);
- Influenza (A or B) positive by rapid antigen test OR a positive RT-PCR test for influenza within the last 24 hours with a CT value of less than 30;
- At least one symptom of influenza (including fever, or history of fever, myalgias, headache, cough, fatigue, nasal congestion, rhinorrhoea, and sore throat) for less than 4 days (96 hours);
- Able to walk unaided and unimpeded in activities of daily living (ADLs);

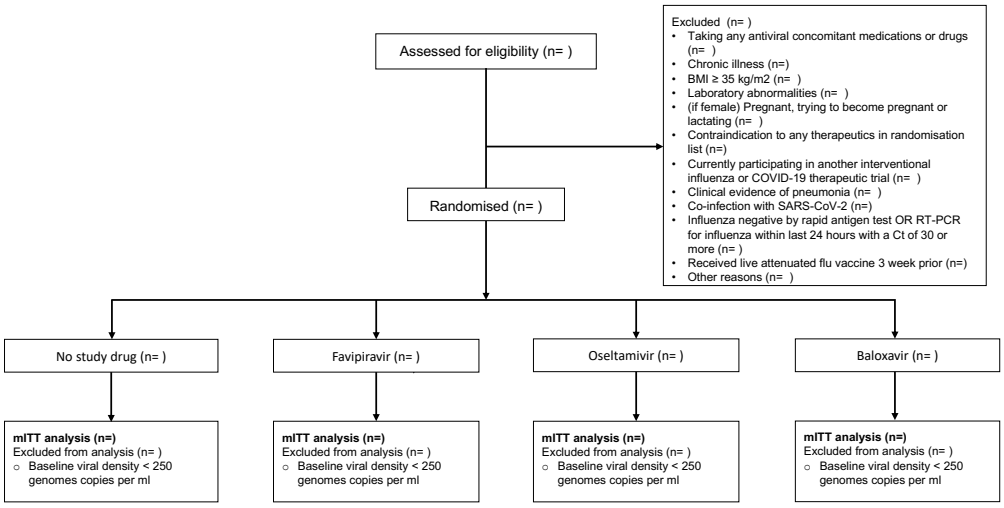

Figure 2: Example CONSORT diagram summarising screening, reasons for excluding patients, randomisation, and number of patients by arm included in the mITT analysis.

5.3 Withdrawal and follow-up

Level of withdrawal will be tabulated and cover the following aspects:

- Discontinuation from any of the study interventions
- Withdrawal from study follow-up
- Withdrawal from entire study and requests that data are not used.

Data will be tabulated for the timing of withdrawal from follow up or lost to follow up. The number of withdrawals and reason will be recorded. A participant may withdraw from the intervention, or be withdrawn for the following reason.

- Withdrawal of consent by participant
- Alteration of participants circumstances or condition which gives justification for discontinuation as decided by investigator.

A table will summarise the loss to follow up and withdrawals during the study with the corresponding reasons.

5.4 Baseline patient characteristics

The following key baseline characteristics will be summarised, stratified by site:

- Age and sex;
- Baseline oropharyngeal viral load (in copies per mL: the mean value of the 4 swabs taken at randomisation);
- Influenza type (A or B)
- Vaccination status (history of vaccinations against influenza within the past 6 months);
- Days since onset of symptoms.

These will be summarised by their mean or median and standard deviation/range for the continuous variables and by their mean for the binary variables.

6 Analysis

6.1 Outcome definitions

The primary outcome is the inferred rate of oropharyngeal viral clearance expressed as a slope coefficient (the mean posterior estimate and the 95% credible interval) for the daily change in the log viral load (on the $\log_{10}$ copies per mL scale). This will be inferred under a Bayesian mixed effects (hierarchical) model using the $\log_{10}$ viral load measurements up until day 5. The observed CT values will be transformed to viral genome copies per mL using the set of control samples from each batch.

6.1.1 Viral load quantification

All viral load quantification will be done using a bespoke ThermoFisher Influenza A and B PCR assay with in-built RNase P. Each qPCR assay is run on a 96 well plate, composed of all the aliquots from the transport media of the swabs from 4 patients (10 timepoints each done in duplicate, i.e. 20 samples per patient) and one control set composed of 6 spiked control samples of known viral densities over the range 10^7 to 10^2 viral copies per mL, each done in duplicate (12 in total). For each plate (i.e. batch) we can thus estimate a standard curve (which defines the conversion from the CT value to a number of viral copies per mL). At each interim analysis we will perform a quality control check by comparing the estimated standard curves from each plate, both within sites and across sites (each country will run the qPCR assays separately). In addition, we will use the standard curve data to estimate heteroscedasticity in the viral load estimation. We expect the qPCR measurement error to increase as the viral load decreases. Heteroscedasticity can be shown visually by plotting the true viral load densities in the control duplicated samples against their observed difference in CT value. We can then estimate the variance in the differences as a function of the true viral density.

To help adjust for variation in the human cell content of the swabbed sample the qPCR method assesses viral densities and also human cell densities using RNaseP as a human cell marker.

6.2 Analysis Methods

6.2.1 Statistical model for the analysis of the serial viral load data

The primary analysis will consist of fitting Bayesian hierarchical (mixed effects) linear models to the $\log_{10}$ oropharyngeal viral load data up until day 5. The treatment effect is defined as the proportional change (expressed as a multiplicative term) in the population slope of the daily change in log viral load. We will model the data on the copies per mL scale, after conversion from CT values using controls from each 96 well plate. The standard curve transformation will be done by fitting a linear mixed effects (random slope and random intercept for each plate with a fixed effect on the intercept for flu type) model to the control data: regressing the CT values on the known log viral densities. This borrows information across plates but allows for batch effects.

The general model likelihood takes the following form:

$$y_{i,t,k} \sim \text{Student}\left(\lambda, \alpha_0 + \alpha_i + \alpha_{cov} + \gamma x_{i,t} + \beta_0 e^{\beta_i + \beta_{cov} + \beta_{T(i)}} t, \sigma^2\right)$$

where:

- $y_{i,t,k}$ is the log viral load ($\log_{10}$ copies per mL) for patient i at time t (in days since randomisation) from site k.
- $T(i)$ is the randomised treatment allocation for individual i.
- σ^2 is the variance of the measurement error.
- λ is the degrees of freedom of the student-t error model.
- The slope of viral clearance decomposes into 5 multiplicative terms: the population mean slope β_0; the individual random effect β_i; the covariate effect β_{cov}; and the treatment effect (fixed across sites and individuals) $\beta_{T(i)}$. All comparisons are made relative to the no antiviral treatment control arm, so we set $\beta_{control} = 0$.

- The intercept term (baseline log viral load) decomposes into 4 terms: the population intercept α_0; the individual random effect α_i; and the sum of any covariate effects α_{cov}.
- $x_{i,t}$ is the relative human RNaseP quantification (RNaseP ΔCT value: this is proportional to the log number of human cells) for patient i at time t. The parameter γ thus provides an adjustment for human cell content in each swab. We scale $x_{i,t}$ so that it has mean 0.

The primary model is the linear model with adjustment for the covariates on the slope and intercept.

We have chosen a student-t distribution for the model likelihood (with the number of degrees of freedom λ estimated from the data) as this is robust against departures from normal (Gaussian) error, and against model mis-specification, notably concerning the assumption of log-linear decline in viral loads. Prior analyses carried out in order to prepare this statistical analysis plan suggest that a considerable number (e.g. up to 5%) of viral load densities can depart substantially from the expected distribution under a simple log-linear model with Gaussian error [4]. This can be explained partially by variation in the recovery from oropharyngeal sampling and partially by the patterns of bi-exponential decay in viral loads. It is unclear if the slope of the second phase of elimination in a bi-exponential decline is affected by interventions. We therefore choose not to fit bi-exponential models as this makes the interpretation of the treatment effect more difficult and prior modelling suggested no major impact in terms of treatment effect estimation [4].

Goodness of fits will be assessed using leave-one-out validation [6] and Bayesian R^2 approximation [7].

6.2.2 Covariate adjustment

In addition to the human RNaseP adjustment, the primary analyses will adjust for the following covariates (through the parameters $\alpha_{cov}, \beta_{cov}$):

- Age (in years - scaled to have mean zero);
- Previous influenza vaccination within the past 6 months (as a binary variable);
- Time since the trial started (in days - scaled to have mean zero). This is to adjust for temporal effects which could confound across arm comparisons made for non-contemporaneously recruiting arms.

If any of the binary covariates show no variation (e.g. all patients are vaccinated) then we will drop the term from the model. We will fit both the covariate adjusted model and the non-adjusted model in all instances as to check for sensitivity to the covariate model. A sensitivity analysis regarding temporal drift will be done by analysing each arm individually along with contemporaneously enrolled controls.

6.2.3 Resolution of fever and symptoms

Time to resolution of fever and time to resolution of symptoms will be analysed by comparing Kaplan-Meier curves (right censored data). Patients are defined as having a fever at baseline if at least one temperature measurement (initially axillary and subsequently aural) within the first 24 hours from randomisation is ≥ 37.5. Switching to aural temperature measurement is under consideration in order to reduce measurement variation. Resolution of fever is defined as an axillary temperature $\leq$37.0C for both measurements taken over one day (at least 24 hours). Resolution of symptoms is defined as no reported symptoms. An additional analysis will look at the proportion of patients with lymphocyte count within the normal ranges at D3 and D7. Differences in time to resolution for both endpoints will be tested using the log-rank test. Differences will be quantified by the change in the median time to resolution.

6.3 Pharmacokinetic analyses

All patients will have 4 plasma samples taken at: randomisation, day 3, day 7 (all three are pre-dose), and finally day 14. Drug concentrations will be quantitated from these samples in order to explore dose-response relationships between drug (or metabolite) exposure and outcome (slope of viral clearance). There are two objectives for the pharmacokinetic data:

- First, to verify that plasma concentrations are within the expected range (using published pharmacokinetic data from equivalent patient groups and dosing);
- Second, to explore whether there is any correlation between plasma concentrations and outcome (rate of viral clearance).

The first objective will use published pharmacokinetic data to assess whether drug levels are within the expected range. For the second objective, the primary analysis will be performed using the day 3 level (a trough level). We will fit a linear model between the day 3 drug (and/or active metabolite) level (log transformed) and the mean viral clearance slope estimate from the main analysis in patients allocated to the intervention arm. Significance is defined as a two-sided test at the 5% level. If appropriate more complex population pharmacokinetic modelling will be performed to reconstruct individual patient concentration profiles in order to relate to the viral clearance rate.

6.4 Subgroup analyses

The primary pre-specified subgroup analysis will examine the relationship between the efficacy of antiviral drugs and the types of influenza (A or B). The subgroup analysis will be carried out only if there is clear in vitro or clinical evidence that there is heterogeneity in effect.

Subgroup analyses will be done by fitting an interaction term $\delta x_{type}(i)e^{\beta_{T(i)}}$ for the relevant $T(i)$; where $x_{type}(i)$ is the subgroup defined by influenza type infecting individual i.

Secondary subgroup analyses will be carried out only for treatment arms that show an antiviral effect (meet the success criteria). We will look at subgroup effects for (listed in order of priority):

- Serological status on admission (using quantitative serum antibody or antibodies);
- Vaccination status;
- Age and sex;

6.5 Prior distributions

In all the following, for the normal distributions, the second term corresponds to the scale (standard deviation not the variance). The viral load values are on the $\log_{10}$ scale (i.e. for the intercept term 6 would correspond to 1,000,000 copies per ml).

Population level parameters

$$\alpha_0 \sim \text{Normal}(6, 2) \quad \text{(population intercept)}$$

$$\beta_0 \sim \text{Normal}(-0.5, 1) \quad \text{(population slope)}$$

$$\sigma \sim \text{Normal}(1.5, 3) \quad \text{(standard deviation of measurement error)}$$

$$\gamma \sim \text{Normal}(0, 1) \quad \text{(human RNaseP adjustment)}$$

$$\beta_{T(i)} \sim \text{Normal}(0, 0.5) \quad \text{log treatment effect (subgroups also)}$$

Covariate effects

$$\beta_{cov} \sim \text{Normal}(0, 1) \quad \text{(covariate effect on slope)}$$

$$\alpha_{cov} \sim \text{Normal}(0, 1) \quad \text{(covariate effect on intercept)}$$

Random effects and hyperparameters

$$\Omega \sim \text{Cholesky}(2) \quad \text{(correlation matrix for individual random effects)}$$

$$\Sigma \sim \text{Exponential}(1) \quad \text{(standard deviation for individual random effects)}$$

$$\lambda \sim \text{Exponential}(1) \quad \text{(t-distribution degrees of freedom)}$$

6.6 Sensitivity analyses

Sensitivity to the prior specification will be assessed by multiplying all prior standard deviation terms by 5.

6.7 Missing data

Missing viral load data will not be imputed. The model will be fit to all available viral load densities up until day 5 or day of treatment deviation. For all analyses, missing data of the key baseline variables (vaccination, virus genotype, antibody rapid test) will be imputed at random using the observed site-specific covariate distribution multiple times (5 times, the model will be refit). We do not expect much missing data (eg <1%).

6.8 Harms

Safety analyses will include all patients who have received at least one dose of the intervention. The safety and tolerability data will be pooled from all the sites that receive the same study intervention. The safety and tolerability will be assessed by comparing the frequency of adverse events and serious adverse events when compared to the control intervention.

All adverse event summaries will refer to treatment emergent adverse events, i.e. adverse events that newly started or increased in intensity after study drug administration (or from hour 0 in the control group receiving no intervention). Adverse events will be graded according to CTCAE V5.0. Adverse Events (AEs) of grade 3 and above will be recorded, the grading is as follows - 1 = mild, 2 = moderate, 3 = severe, 4 = life-threatening, 5 = fatal. All Serious Adverse Events (SAEs) will be recorded, a serious adverse event is any untoward medical occurrence that results in death, is life threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant disability/incapacity or consists of a congenital anomaly or birth defect.

We will tabulate AEs of grade 3 and higher, SAEs, and deaths, by treatment arm with comparisons made relative to the no study drug arm.

None of the drugs evaluated will be pre-registration or pre-approval, so adverse effect profiles for each are already well known.

6.9 Statistical software and analysis implementation

All analyses will be be in done in R. Interim analyses and the final analyses will be done using RMarkdown scripts, pre-coded to ensure full transparency. All code for the statistical analysis is available on the following github repository: https://github.com/jwatowatson/ADASTRA-SAP. This ensures full version control by tracking all changes made to the analysis code.

The statistical models for the analysis of the serial viral load data are written in *stan* and fitted via the *rstan* interface [8]. These are available on the github repository. For each model, we will run 6 parallel chains for 10,000 iterations, discarding half for burn-in and thinning every 40 (thus 1,000 posterior draws).

Convergence of Bayesian fits will be assessed visually by examining the traceplots of the key model parameters and assessing the Rhat values for each parameter (Gelman-Rubin statistic).

References

[1] Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343. doi:10.1001/jama.2017.18556.

[2] Schilling WH, Jittamala P, Watson JA, Ekkapongpisit M, Siripoon T, Ngamprasertchai T, et al. Pharmacometrics of high dose ivermectin in early COVID-19: an open label, randomized, controlled adaptive platform trial (PLATCOV). ELife. 2023;12.

[3] Hayden FG, Sugaya N, Hirotsu N, Lee N, de Jong MD, Hurt AC, et al. Baloxavir Marboxil for Uncomplicated Influenza in Adults and Adolescents. New England Journal of Medicine. 2018;379(10):913–923. doi:10.1056/NEJMoa1716197.

[4] Watson JA, Kissler SM, Day NPJ, Grad YH, White NJ. Characterizing SARS-CoV-2 Viral Clearance Kinetics to Improve the Design of Antiviral Pharmacometric Studies. Antimicrobial Agents and Chemotherapy. 2022;66(7):e00192–22. doi:10.1128/aac.00192-22.

[5] Wongnak P, Schilling WH, Jittamala P, Boyd S, Luvira V, Siripoon T, et al. Temporal changes in SARS-CoV-2 clearance kinetics and the optimal design of antiviral pharmacodynamic studies: an individual patient data meta-analysis of a randomised, controlled, adaptive platform study (PLATCOV). The Lancet Infectious Diseases. 2024;.

[6] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing. 2017;27(5):1413–1432.

[7] Gelman A, Goodrich B, Gabry J, Vehtari A. R-squared for Bayesian Regression Models. The American Statistician. 2019;73(3):307–309. doi:10.1080/00031305.2018.1549100.

[8] Stan Development Team. RStan: the R interface to Stan; 2020. Available from: http://mc-stan.org/.

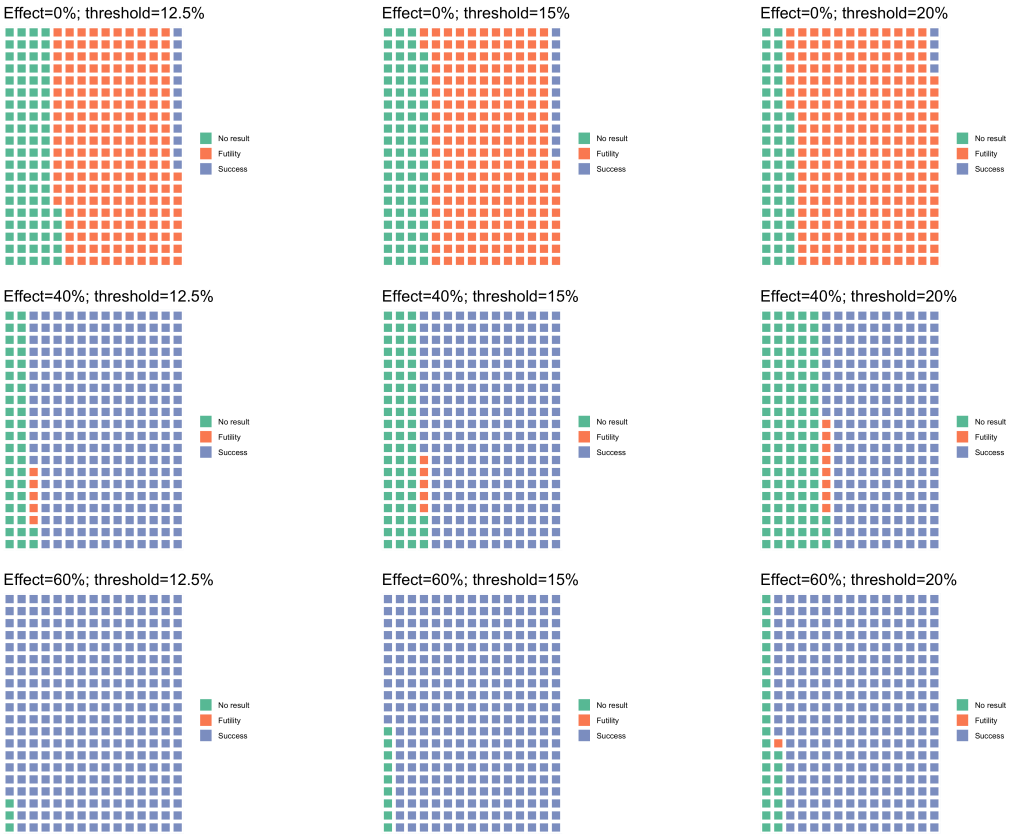

Figure S1: Waffle plots showing the proportion of each outcome (success: blue; futility: orange; no result by 120 patients: green) for the success versus futility stopping rule. 100 simulations were run for each permutation of the effect size and λ_1 threshold.

S1 Updated sample size simulations

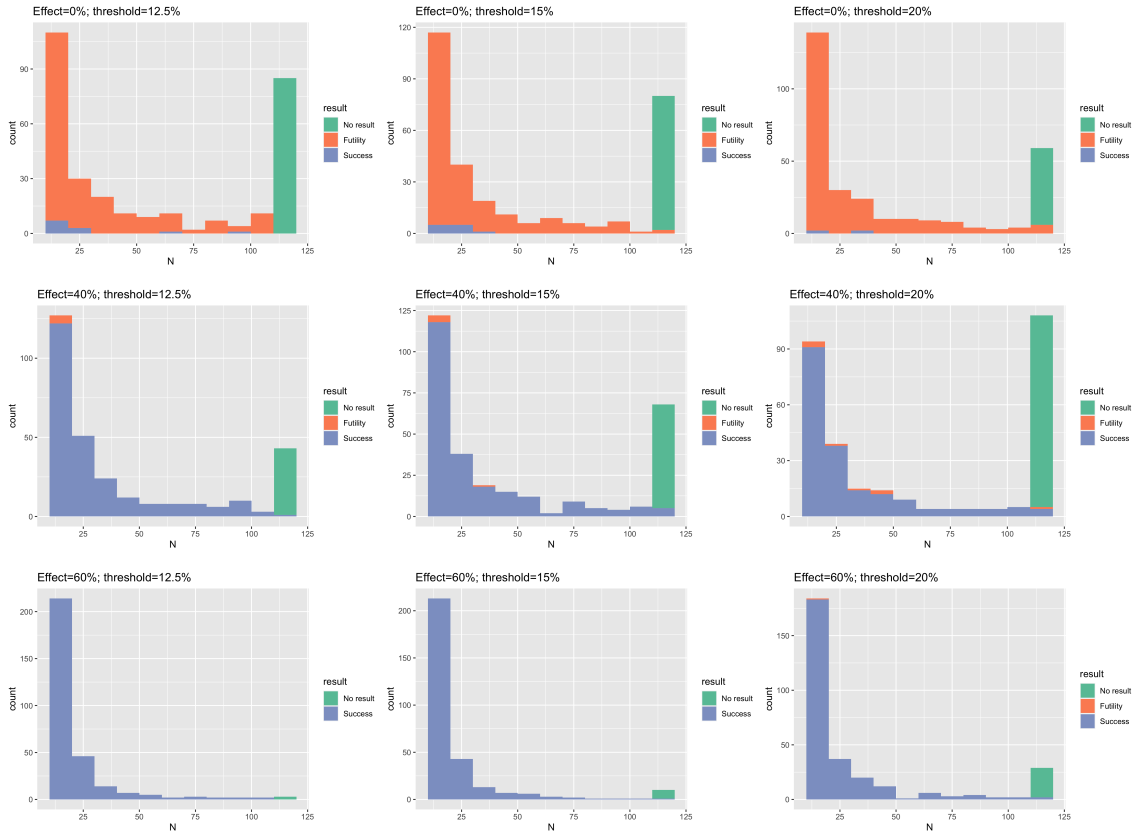

Figure S2: Histogram of sample size until stopping rule is met, same colors as in Figure S1.

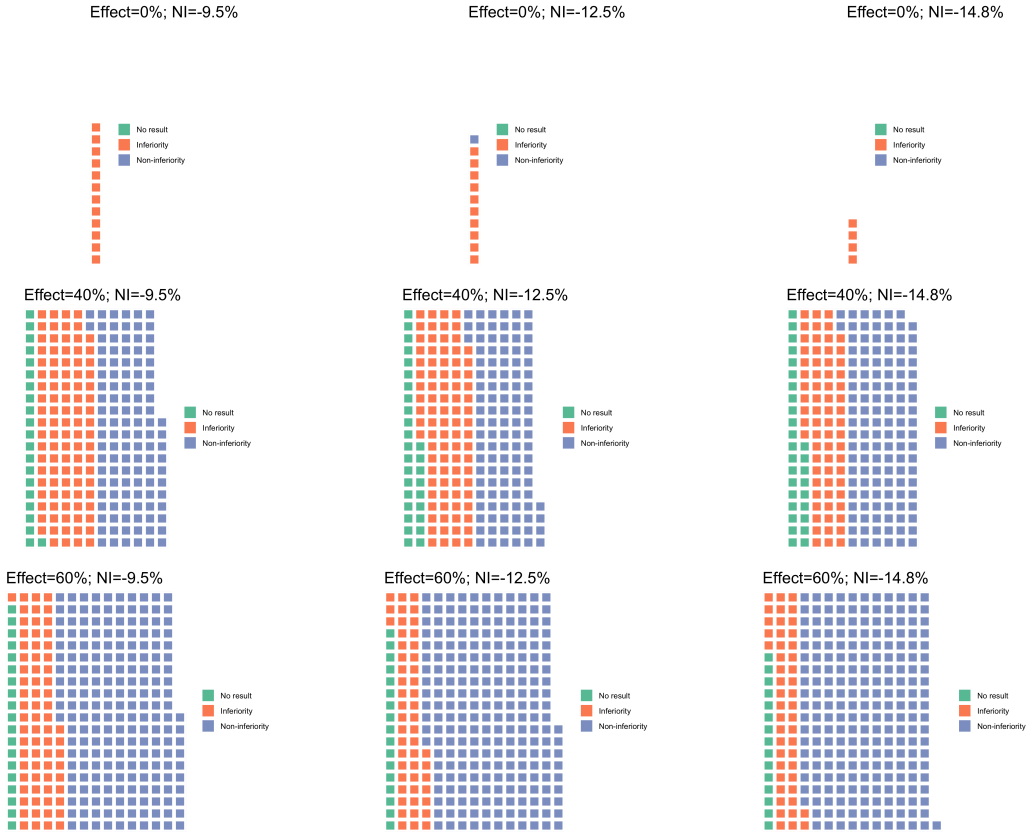

Figure S3: Waffle plots showing the proportion of each outcome (non-inferiority: blue; inferiority: orange; no result by 120 patients: green) for the non-inferiority stopping rule. 100 simulations were run for each permutation of the effect size and λ_2 threshold.

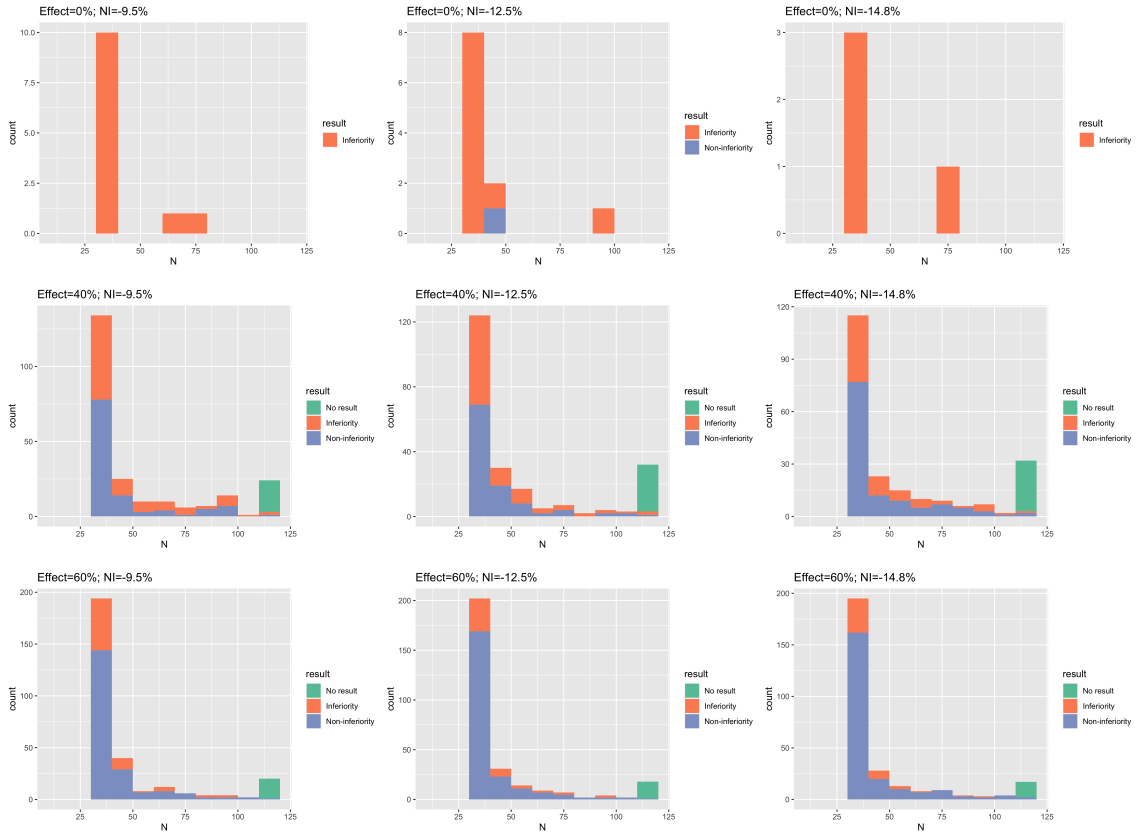

Figure S4: Histogram of sample size until stopping rule is met, same colours as in Figure S3.