CCTU/TPL007V3 Approved:24/01/2018 Reviewed 24/01/2018

Statistical Analysis Plan

TRIAL FULL TITLE	**Impact of increased praziquantel frequency on childhood fibrosis in persistent schistosomiasis morbidity hotspots.**
EUDRACT NUMBER	
SAP VERSION	V3.0
ISRCTN NUMBER	16994599
SAP VERSION DATE	05/06/2024
TRIAL STATISTICIAN	Dr Simon Bond
TRIAL CHIEF INVESTIGATOR	Dr Shona Wilson
SAP AUTHOR	Rachel Donegan

1 SAP Signatures

I give my approval for the attached SAP entitled FibroScHot Final SAP V3.0 dated 05/06/2024

Chief Investigator

Name: Dr Shona Wilson

Signature: Shona Wilson Date: 11/6/24

Statistician

Name: Dr Simon Bond

Signature: _____ Date: _____

2 Table of Contents

3 Abbreviations and Definitions

AE	Adverse Event
CACE	Complier Average Causal Effect
CCTU	Cambridge Clinical Trials Unit
CCG	Circulating Cathode Antigen
CI	Confidence Interval
CRF	Case Report Form
EMA	European Medicines Agency
EPG	Eggs Per Gram
ERR	Egg Reduction Rate
FDA	Food and Drug Administration
IMP	Investigational Medical Product
MDA	Mass Drug Administration
NTD	Neglected Tropical Disease
PC	Preventative Chemotherapy
PZQ	Praziquantel
SAP	Statistical Analysis Plan
SCORE	Schistosomiasis Consortium for Operational Research and Evaluation

SAE / SAR	Serious Adverse Event /Serious Adverse Reaction

4 Introduction

4.1 Preface

Periportal fibrosis is the severest pathology caused by *Schistosomiasis mansoni* (1). The cornerstone of disease control is annual preventative chemotherapy (PC) through mass drug administration (MDA) of the safe and efficacious drug praziquantel (PZQ) (2). However, in areas of intense transmission ('hotspots') of the disease, MDA is failing to decrease infection intensity (3–5). Worryingly fibrotic patterns are not restricted to those of the older generation, ruling out the scenario that the fibrosis present is due to childhood exposure amongst those too old to have received treatment through MDA when they were school aged. This trial will investigate the effectiveness of increasing the number of treatment doses of PZQ given throughout the year in an infection hotspot on the prevalence of childhood periportal fibrosis. The standard annual dosing will be compared with two intervention arms, one with twice yearly dosing and one with four times yearly dosing.

4.2 Purpose of the analyses

The analysis will evaluate whether increasing treatment frequency with praziquantel through school- based /school-community based delivery strategies reduces the prevalence of childhood periportal fibrosis in schistosomiasis morbidity hot-spots after two years of intervention, comparing the standard one dose per year to two intervention arms, one with a twice yearly dosing and the other with a four yearly dosing.

5 Study Objectives and Endpoints

5.1 Study Objectives

(ICH E3; 8)

5.1.1 Primary objective

To compare the effect of 2x and 4x treatments annually with praziquantel to the standard 1x annual treatment on the prevalence of *Schistosoma mansoni* associated periportal fibrosis of the liver.

5.1.2 Secondary objective

To compare the effect of 2x and 4x treatments annually with praziquantel to the standard 1x annual treatment on the mean infection intensity of *Schistosoma mansoni*.

5.1.3 Exploratory objectives

Compare the effect of 2x and 4 x treatments annually with PZQ to the standard 1x annual treatment on the Egg Reduction Rate (ERR)

Compare the effect of 2x and 4 x treatments annually with PZQ to the standard 1x annual treatment on the point prevalence.

Compare the effect of 2x and 4 x treatments annually with PZQ to the standard 1x annual treatment on the cure rate.

Estimate the efficiency of using infection intensity to predict the occurrence of AE's.

5.2 Endpoints

(ICH E9: 2.2.2)

5.2.1 Primary endpoint

- Periportal fibrosis measured at 2-years post 1st treatment follow up, defined as Niamey Protocol pattern scores of C, D, E or F = 1 v's A, B = 0 (scores X, Y & Z will not be analysed see section 8.4)

5.2.1.1 *Secondary endpoint*

- *Schistosoma mansoni* infection intensities as measured by Kato-Katz (KK) faecal egg counts (6), measured at 2-years post 1st treatment follow up

5.2.2 Exploratory endpoints

- *Schistosoma mansoni* Egg Reduction Rate (ERR) measured between:
    - Point Baseline and baseline +4 weeks
    - Baseline and 2 years + 4 weeks
    - 2 years and 2 years +4 weeks
- Prevalence at 2 years post 1st treatment
- Cure rate between:
    - Baseline and baseline +4 weeks
    - Baseline and 24 months
    - Baseline and 24 months +4 weeks
- Adverse Events measured at baseline and 24 months:
    - Any AE (Yes/No)
    - Digestive AE's - 4,5,6,7,or 12 on CRF see Also section 11 of SAP (Yes / No)
    - Neurological AE's - 1,2,3,8,9 or 11 on CRF see Also section 11 of SAP (Yes/No)

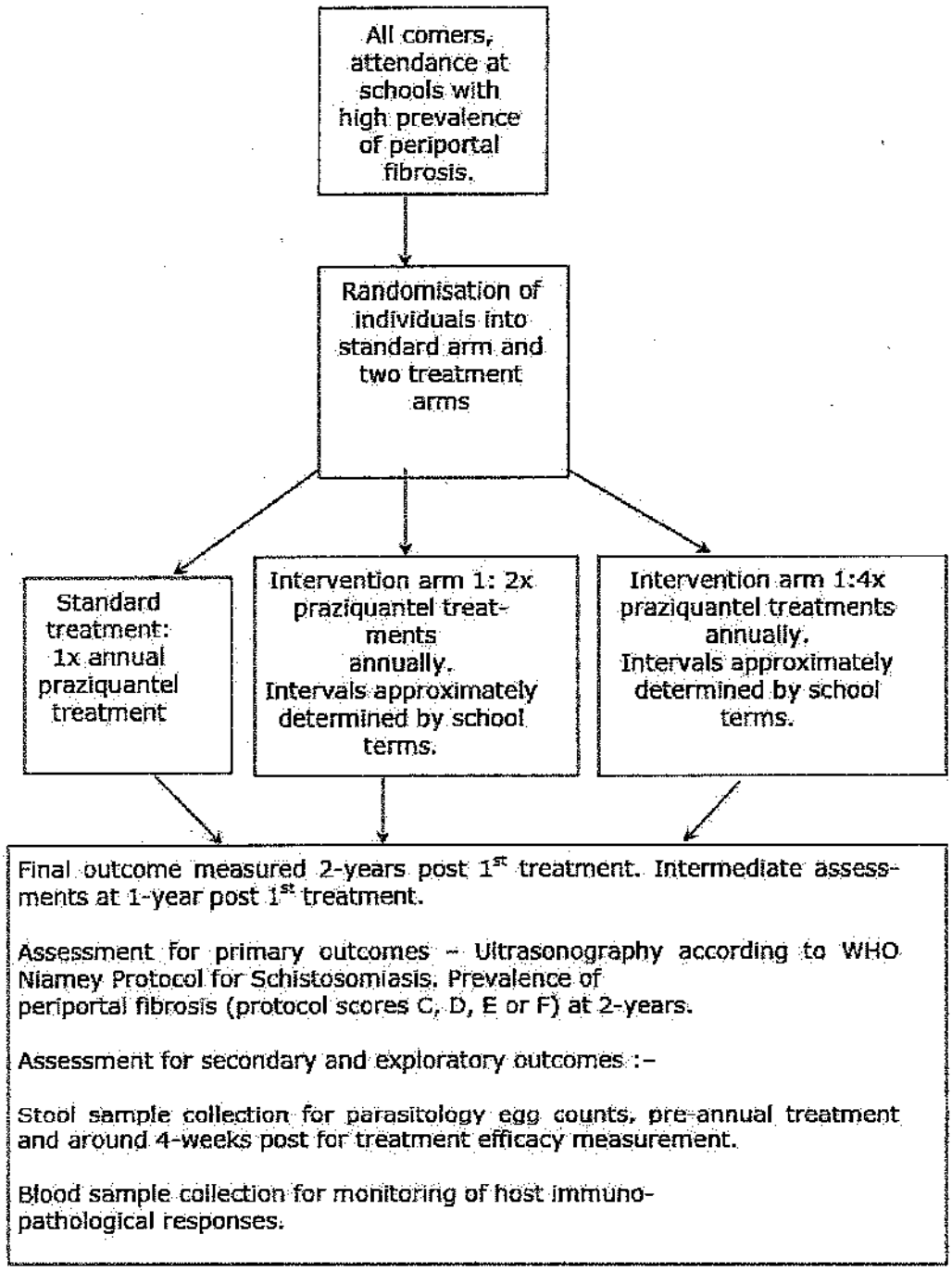
6 Study Methods

6.1 General Study Design and Plan

(ICH E3; 9)

This is a phase IV, randomised superiority trial that is blinded to the primary outcome assessors, but open label to all other investigators and participants. The control arm is the standard annual dose of praziquantel, intervention arm 1 is a twice yearly dose of praziquantel and intervention arm 2 is a four yearly dose of praziquantel. Randomisation will be 1:1:1 to each arm and stratified by school.

Trial Flow Chart



NB! Only 120 participants sampled at 12 months and only parasitology egg counts and only pre-treatment.

6.2 Inclusion-Exclusion Criteria and General Study Population

(ICH E3; 9.3. ICH E9; 2.2.1)

6.2.1 Inclusion Criteria

To be included in the trial the participant must:

- Have a parent/guardian who has given written/marked informed consent for the child to participate
- Have given written/marked informed assent to participate
- Be aged 6–14–years of age.
- Be born or resident within the community in which their school is situated for >=2 years.

6.2.2 Exclusion Criteria

The presence of any of the following will preclude participant inclusion:

- History of facial oedema after treatment with praziquantel
- Known neurocysticercosis or intraocular cysticercosis

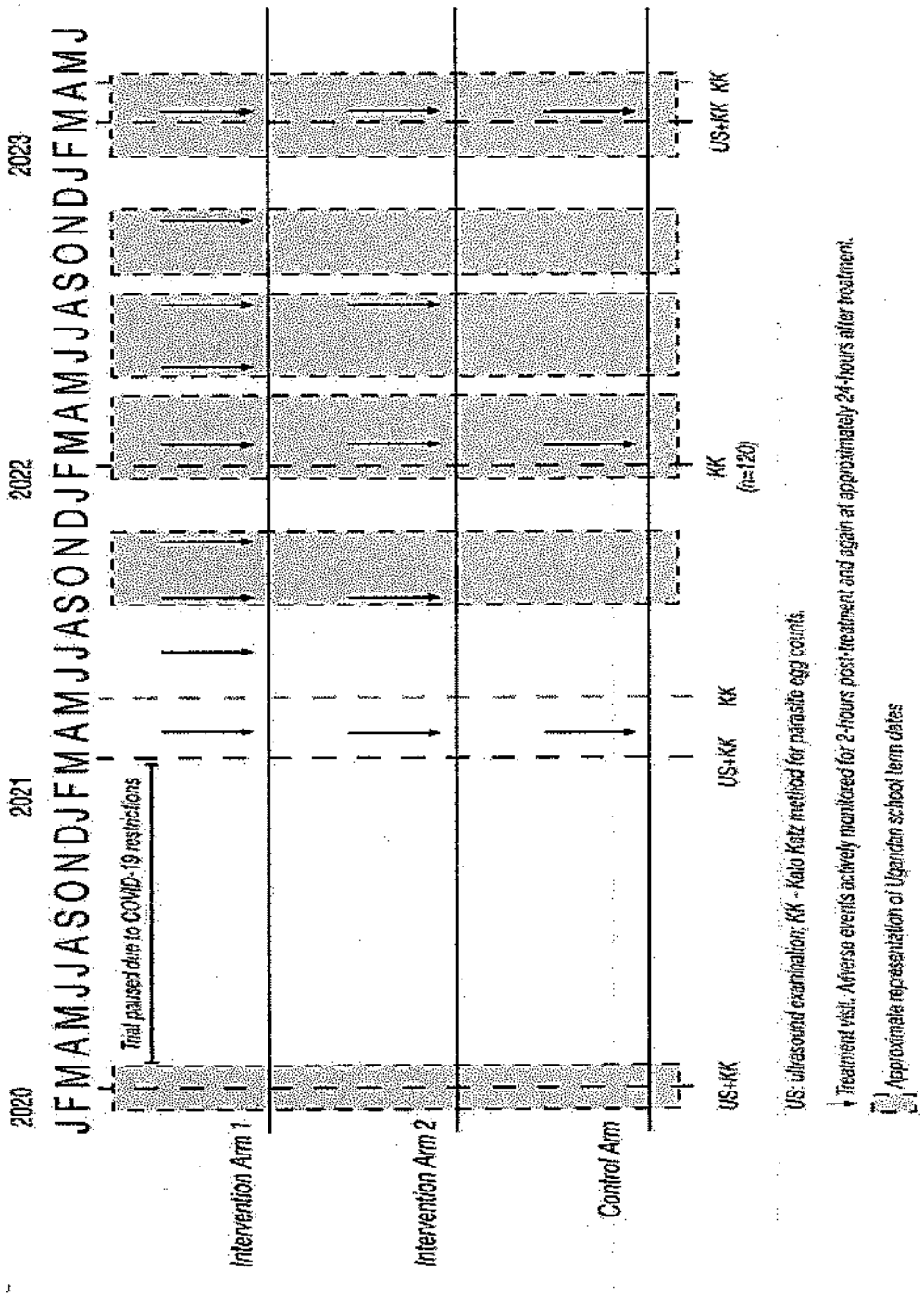
6.3 Randomisation and Blinding

(ICH E3; 9.4.3, 9.4.6. ICH E9; 2.3.1, 2.3.2)

Prior to enrolment trial participant identification numbers will be generated and these numbers assigned an intervention arm via simple computer generated random number randomisation. This paper-based list will be provided to the trial site team and upon enrolment children will be assigned to the ordered participant identification numbers, class-by- class and school-by-school and school-by-school or household-by- household if schools are not open. Randomisation will be stratified by school. The trial is open label to participants and the investigators, with the exception of the ultra-sonographers who are undertaking the scoring for the primary trial outcome. Scoring will be undertaken remotely from the participant being examined and the technicians conducting the exam.

6.4 Study Variables

(ICH E3; 9.5.1. ICH E9; 2.2.2)



- WHO ultrasonography Niamey protocol pattern scores – are a means of measuring periportal fibrosis. Ultrasonographic images of the liver parenchyma are compared with standard image patterns of liver parenchyma using annex A of 'A Practical Guide to the Standardized Use of Ultrasonography for the Assessment of Schistosomiasis-related Morbidity' and a score assigned as detailed below:

Pattern	Picture	IP-Score
A	Normal structure	0
Patterns observed in schistosomiasis		
B	'Starry sky' (diffuse echogenic foci)	1
C	Highly echogenic 'ring echoes', which correspond to the 'pipe stems' seen in a scan perpendicular to the one where rings are seen	2
D	Highly echogenic 'ruff' around portal bifurcation and main stem	4
E	Highly echogenic 'patches' extending* expanding? from the main portal vein and branches into the parenchyma	6
F	Highly echogenic 'bands' and 'streaks' , extending from the main portal vein and its bifurcation to the liver surface, where they retract the organ surface. (Note: echogenic thickening of ligamenta alone does not justify assignment of the sonographic image to this pattern.)	8
X	Diffusely coarse liver texture, irregular liver surface, distorted hepatic veins, rounded caudal liver edge	-
Y	Diffusely increased liver echogenicity, loss of highly reflective edges of peripheral portal branches, possibly distal sound extinction, rounded caudal liver edge	-
Z	All other liver abnormalities, specify	-

Additional category created to accommodate participants who have a real and recorded fibrosis pattern of A|B OR B|A. Since there is no adjudication for these participants they will be recorded in summary statistic tables as A/B. This will not however have any impact on the primary endpoint since these individuals will still be classed as not having peri-portal fibrosis for the binary variable.

Ultrasonography will be conducted at the following time points:

- Baseline
- 24 months

Kato-Katz egg counts - this is a faecal thick microscopy smear used to calculate the *Schistosoma mansoni* eggs per gram of faeces. Up to three stool samples are collected at each of the following time points:

- Baseline
- Baseline + 4 weeks
- 12 months
- 24 months
- 24 months + 4 weeks

A maximum of six KK slides will be prepared from the three stool samples and the final egg counts recorded on the parasitology CRF for each visit are calculated as the mean of the total KK slides for that visit.

7 Sample Size

(ICH E3; 9.7.2. ICH E9; 3.5)

Berhe et al 2008 (7) reported a 47.7% resolution rate of grade C to non-periportal fibrosis scores of A and B and 24.6% resolution rates of D scores to scores A and B. Extrapolating this success rate onto our preliminary data, we would expect 35% less of the children in a successful intervention arm to have periportal fibrosis after 2years intervention than those in the control arm. However, Berhe et al also reported geographical differences in resolution, with higher transmission villages having poorer resolution. We therefore propose a minimum 20% absolute difference in prevalence of periportal fibrosis between the control and intervention arms to be a clinically significant outcome. Based on preliminary data we assume an incidence rate of scores C-F of 80% in the control arm.

Initial sample size calculations used an adjusted alpha to allow for multiple testing using the Bonferoni correction - " .... a Bonferroni correction to allow any pairwise comparison between the three arms, testing at a 1.66% significance level, a total sample size of 450, 150 per arm, would provide 92% power to detect an absolute reduction of 20%." In previous Lake Victoria fishing community observation studies, failure to follow-up has been as high as 35%. The communities on Lake Albert are more stable (NB Kabatereine, pers. comm.) but to minimise failure to follow-up, selection was targeted at children in P1 to P4, retaining the cohort within the primary school structure over the two years of the study, and an inclusion criteria of born or resident in the village for >2yrs, displaying stability of their household within the village is included. Thus a lower estimate of failure to follow-up of 25% was included in the final study numbers giving a final sample size of 600.

The impact of COVID-19 on the study sample size however was a concern, some participants enrolled in the study were not located once the study had re-started as many had relocated due to the pandemic. Some "catch-up" enrolment beyond the initially aimed for 600 was conducted, however, tracing of the participants continued to be compounded by the fact that IMP administration was now happening in the community as schools had closed, and severe flooding and impacts of illegal fishing gear destruction by the government on parental livelihoods resulted in further community instability.

Also during the monitoring conducted by CCTU on behalf of the sponsor around 25% of participants in one of the primary schools had missing consent documentation. Where participants cannot be re-consented they will be withdrawn from the trial which may have an impact on the power of the study if the numbers are significant.

A gate-keeping approach to multiple testing was outlined in the Abstract SAP V1.0 and section 8.6 below, which allowed a significance level (alpha) of 5 % to be used as it is completely apportioned to each hypothesis in sequence and thus only used once and only if the significance criteria is met. A reduction of 20 % with 100 participants per arm would provide a power of 88%, a reduction of 25% would give 97% power and a 30% reduction would give 99% power.

8 General Considerations

8.1 Timing of Analyses

The final analysis will occur after the 24 month + 4 week parasitology visits have taken place and all data from this visit has been entered into the macro database.

Version 1.0 of the SAP will be approved according to the requirements of CCTU/SOP023 before the data download. Permission to lock the database (softlock) will be requested by the chief investigator (CI) using CCTU/FRM094, and the data programmers will deposit the data in a folder to which the trial statistician is given access. Logical and graphical checks will be carried out to check that no out of range values are present. Data queries will be referred to the data management team for resolution (using CCTU/TPL064). Once all data queries have been resolved form CCTU/FRM094 will be signed off for database hardlock approval and at this point the final analysis report produced.

8.2 Analysis Populations

(ICH E3; 9.7.1, 11.4.2.5. ICH E9; 5.2)

8.2.1 Full Analysis Population

- *All participants who were consented, eligible and randomised*
- *Any participant who had an ultrasound score of X,Y or Z at baseline will be withdrawn from the trial and is therefore not included in the 'Full Analysis Population'*
- *Any participants whose re-consent was required but not obtained are withdrawn from the trial and are not therefore included in the 'Full Analysis Population'*

8.2.2 Safety Population

- *All subjects who were consented and randomised and received any dose of the IMP irrespective of whether or not they were eligible.*

8.3 Covariates and Subgroups

(ICH E3; 9.7.1, 11.4.2.1. ICH E9; 5.7)

School site was used to stratify randomisation and will be adjusted for as a fixed coefficient in the primary analysis.

Baseline Fibrosis score and egg count will be fixed effects in the model.

8.4 Missing Data

(ICH E3; 9.7.1, 11.4.2.2. ICH E9; 5.3. EMA Guideline on Missing Data in Confirmatory Clinical Trials)

The initial sample size calculation allowed for a 25% loss to follow up. However the power of the study has since been estimated based on a re-estimated prediction of the final sample size and a revised approach to multiple testing. If therefore the primary outcome data does not fall below 300 participants, then a complete case analysis of the data will be carried out. A graph to illustrate the proportion of the missing data for the primary and secondary endpoints as well as the baseline variables of periportal fibrosis and infection intensity (EPG) will be produced.

Data will be assumed to be missing at random (MAR) and appropriate sensitivity analysis conducted to assess the treatment effect for the primary analysis (8).

8.4.1 X, Y and Z scores for the primary efficacy analysis

The WHO ultrasonography Niamey protocol pattern scores include X, Y and Z which are not analysed in this trial as the liver patterns observed for these scores mean that it difficult to tell if there is periportal fibrosis due to Schistosoma.

At baseline any participant with these scores will be marked as 'unclassifiable and removed from cohort' they do not contribute any further data to the trial. They will be excluded from the Full Analysis Population.
Any participant with Ultrasonography scores of X, Y or Z at 2 years will be excluded from any analysis of the primary endpoint as missing data. This is because the number of participants is likely to be small, around 1% of the sample size.

Participants with X, Y and Z scores for the primary end point could however contribute data to the Secondary and exploratory endpoints.

8.4.2 Missing baseline data

There is a proportion of missing baseline WHO ultrasonography Niamey protocol pattern scores and egg count data. Only using cases with completed baseline data for these variables will significantly reduce the efficiency of the model as it may exclude outcome data which has been collected.

We will assume that whether the baseline data is collected or not is not associated with the treatment assigned. Missing baseline values will be calculated using predictive mean matching (egg count) or logistic regression (fibrosis score) imputation using the "mice" package in R . (9,10)

8.5 Interim Analyses and Data Monitoring

(ICH E3; 9.7.1, 11.4.2.3. ICH E9; 4.1, FDA Feb 2010 "Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics")

No interim analysis will be required for this trial.

8.6 Multiple Testing

(ICH E3; 9.7.1, 11.4.2.5. ICH E9; 2.2.5)

To allow for multiple testing of the primary endpoint and control for the type I error rate (false positive error rate) a gatekeeping procedure will be used. The null hypothesis of no difference in the prevalence of periportal fibrosis will be tested for each of the following comparisons and compared to a significance level (α - alpha) of 5%:

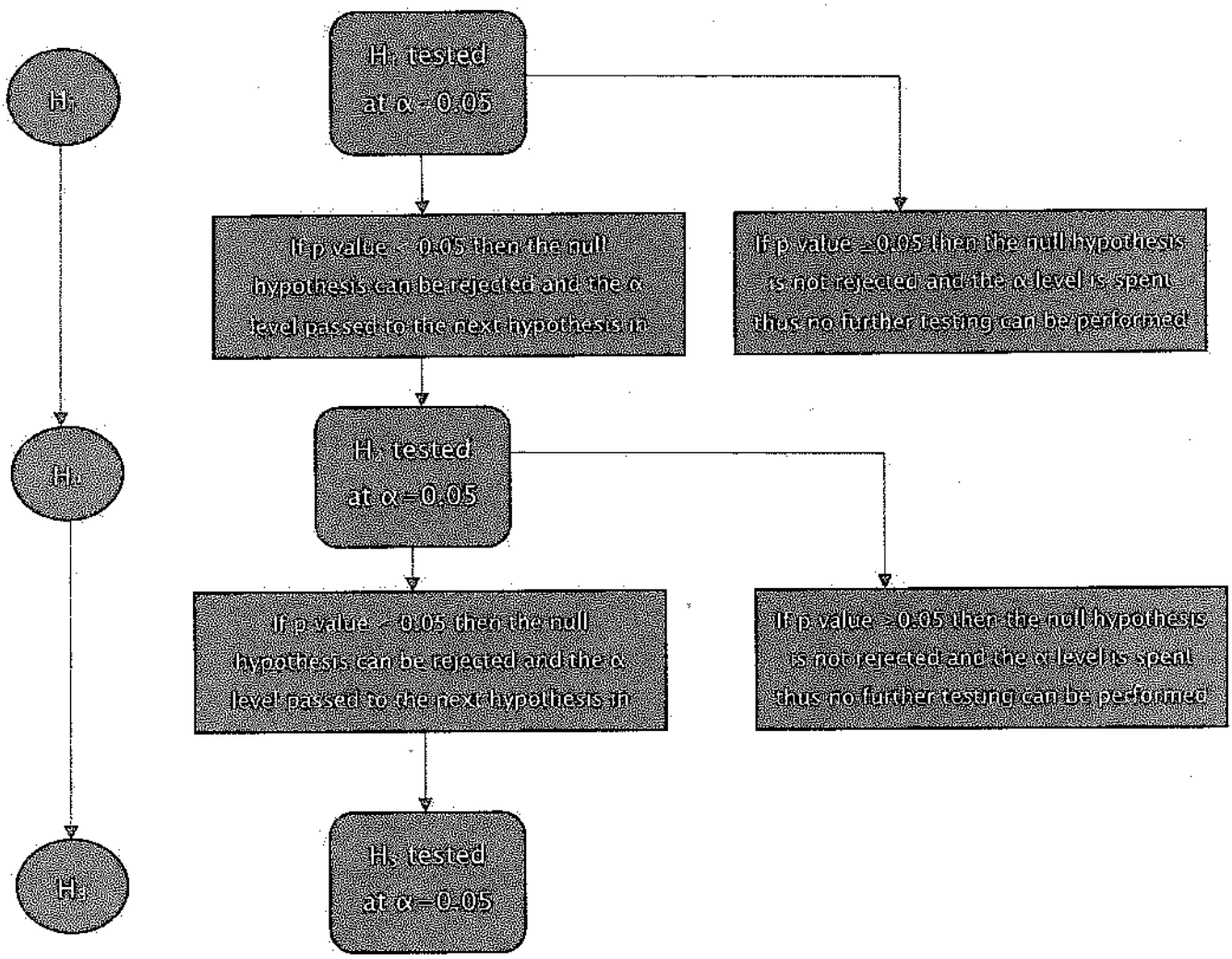
H_1: - 4 x annual treatment of PZQ v's Standard 1x annual treatment

H_2: - 2 x annual treatment of PZQ v's Standard 1x annual treatment

H_3: - 4 x annual treatment of PZQ v's 2 x annual treatment

Each hypotheses will be tested ***in sequence*** and testing continues until the null hypothesis cannot be rejected i.e. a non-significant result is obtained (Fig 2), or all three comparisons are statistically significant. The sequential testing rules mean that the overall rate of any false positive across the three possible comparisons is constrained to be less than 5% (11).

Fig 2: Flow diagram to illustrate testing of hypotheses using gate-keeping procedure



This is a deviation from the multiple testing approach detailed in the protocol v 2.0. The justification is to use the ordering implied by the details of the dosing: so if the most frequent dosing cannot show an improvement in efficacy compared to the least frequent, then we can assume the intermediate dosing will not do so either, and there would be no value in carrying out formal inference. This approach will help to maintain the power of the most relevant questions, given the anticipated reduction in sample size due to movement of the population as a result of the COVID 19 pandemic and other impactful local conditions.

9 Summary of Study Data

All continuous variables will be summarised using the following descriptive statistics: n (non-missing sample size), mean, standard deviation, median, maximum and minimum. The frequency and percentages (based on the non-missing sample size) of observed levels will be reported for all categorical

measures. In general, all data will be listed, sorted by treatment and subject, and when appropriate by visit number within subject. All summary tables will be structured with a column for each treatment in the order (Standard of Care, Intervention 1 (2 x treatment), Intervention 2 (4 x treatment) / Total) and will be annotated with the total population size relevant to that table/treatment, including any missing observations.

The sample size of non-missing values for univariable summary statistics may be larger than the sample size used in the primary regression analysis.

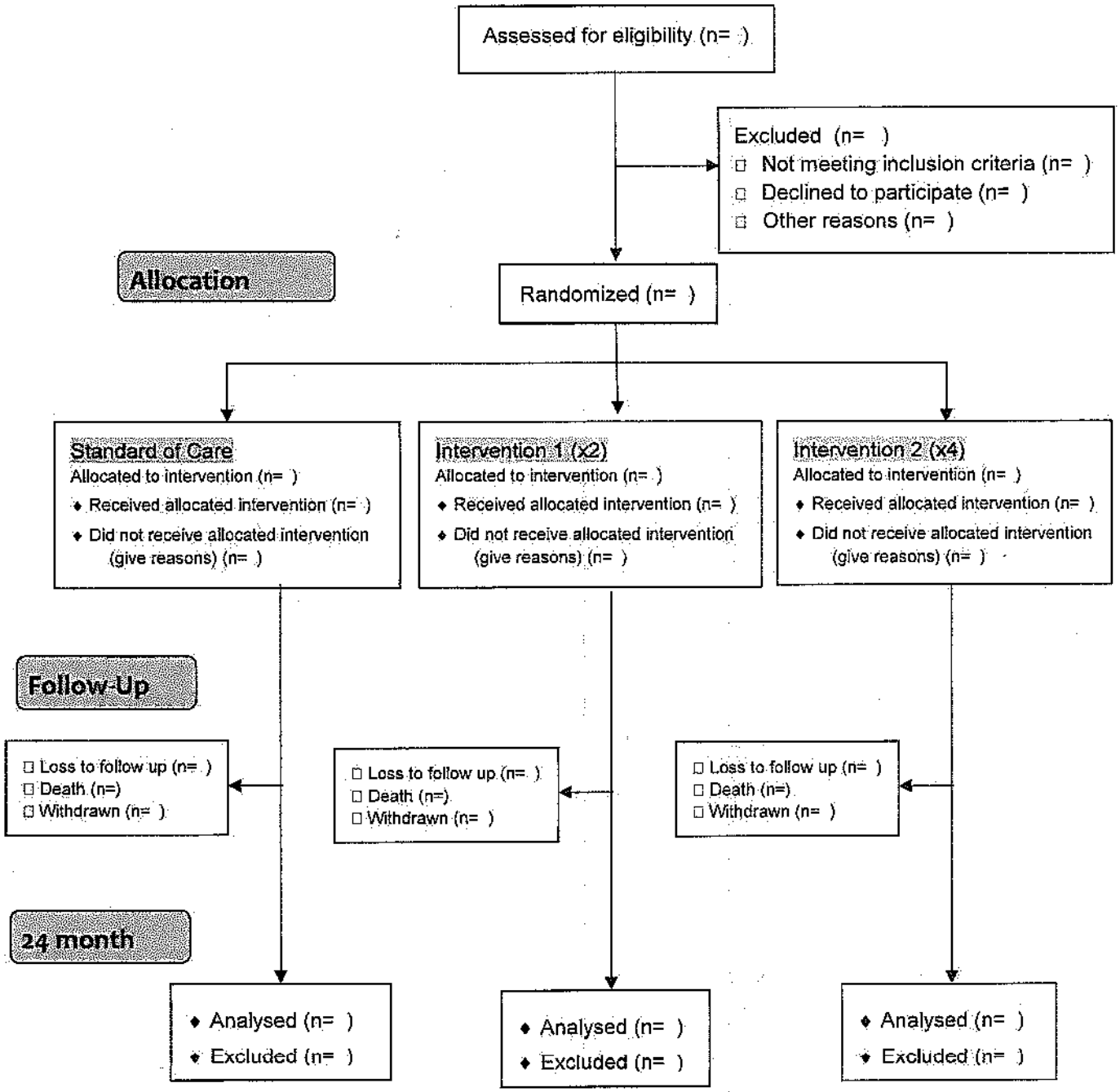
9.1 Subject Disposition

9.1.1 Subject Disposition

Patient disposition of the number of participants who consent, who are eligible and who are randomised as well as drop-out rates will be determined using information collected from the following CRF forms:

- Number assessed for eligibility: Inclusion criteria CRF (check all questions completed) / Exclusion criteria CRF (check all questions completed) & Confirmation of eligibility CRF (eligibility criteria confirmed)
- Number eligible /not eligible: Inclusion criteria CRF / Exclusion criteria CRF & Confirmation of eligibility CRF
- Number randomised: Randomisation CRF
- Number randomised to treatment arm who received treatment allocation: Randomisation CRF, IMP administration for all visits,
- Number for whom primary analysis data collected at 24 months: Ultrasonography CRF for 24 month visit
- Number withdrawn (including deaths) between Baseline and 24 months: Withdrawal CRF (withdrawal data / reason for withdrawal), End of study visit (end of study date /reason for termination / last study visit)

9.1.2 Consort Flow Diagram



9.2 Derived variables

- Periportal Fibrosis is a binary outcome:

WHO ultrasonography Niamey protocol pattern scores of C, D, E or F = 1 (i.e. has periportal fibrosis)

WHO ultrasonography Niamey protocol pattern scores of A or B = 0 (i.e. does not have periportal fibrosis)

- Kato-Katz egg counts
  - Eggs Per Gram

The epg value for each participant visit is calculated in the macro database as the mean of the kato-katz slide egg counts from a maximum of 3 stool samples multiplied by an adjustment factor (one stool sample is sufficient for protocol adherence at a particular time point).

Helminth distributions within the human host are highly over dispersed therefore the epg values for each participant will be log transformed using natural logs (ln).

A constant value of 1 will be added to all raw epg values to allow for the possibility of an epg value of 0.

$$\sum(\ln(epg + 1))$$

The geometric mean epg will be calculated as below:

$$e^{\frac{\Sigma(\ln(epg+1))}{n}} - 1$$

95% CI's calculated on the logarithmic scale can also be converted back to the original scale by exponentiating and subtracting 1.

  - Egg Reduction Rate (ERR)

Sample Egg Reduction Rate will be estimated as:

$$(1-(epg2 \div epg1) \times 100)$$

Where epg1 and epg2 are the arithmetic mean of epg of faeces at:

- baseline and at 24 months + 4 weeks
- baseline and baseline + 4 weeks
- 24 months and 24 months +4 weeks

- Heavy Infection

The WHO classifies 'Heavy infection' as those with epg counts of ≥400. Where the percentage of children with of heavy infection is < 5% then morbidity control is considered to have been achieved and if the percentage of heavy infection is < 1% then the disease is considered to have been eliminated as a public health concern.

Heavy Infection: Binary variable - 1 = "yes" = epg ≥400, 0 = "no" =epg < 400 or not detectable

- Disease status

Disease status is a binary variable where 0 = No disease (epg =0); 1 = Disease (>0 epg)

Point Prevalence (PP)

Is a measure of disease frequency and quantifies the proportion of individuals in a population who have Schistosomaisis mansoni at a specific instant in time.

$$PP = \frac{Total\ number\ of\ participants\ who\ have\ an\ average\ epg > 0}{\text{Total number of participant at risk}}$$

Total number of participants at risk will include all participants for whom at least one stool sample was collected at the timepoint being assessed.

Cure Rate (CR)

The CR is the proportion of participants who have changed from an epg >0 to and epg of = 0 between timepoints;

$$CR = \frac{Total\ number\ of\ participants\ who\ have\ an\ average\ epg = 0\ at\ timepoint\ 1}{\text{Number of participants who have epg} > 0 \text{ at timepoint } 0}$$

Residency in Hoima district is dichotomised into 2 years - 4 years/ > 4 years

9.3 Protocol Deviations

All school-based trial activities were paused on 19th March 2020 in accordance with the Ugandan government announcement concerning COVID-19. Therefore, baseline data for the trial had been collected at different time-points.

Baseline data collected from Buhirigi was collected 17 months after MDA, in March 2020, Baseline data for Kaiso was collected 9 months after MDA, in March 2021.

In order to ensure data integrity a 10% resample of the Buhirigi baseline data was performed in March 2021 for the primary and secondary endpoints.

9.3.1 Baseline data Unpaired Analysis

Data for participants who had baseline data re-sampled in March 2021could not be linked to their baseline data taken in March 2020 since ID data required to match the participants was not recorded / not available. Therefore an unpaired analysis as detailed below will be performed since the matched analysis detailed in the Final SAP V1.0 section 9.3 is not possible due to lack of linkage information.

9.3.2 Schistosoma mansoni infection intensity by faecal egg count

An Empirical cumulative distribution plot of ln(epg+1) of all 2020 and the 2021 resample will be produced. The Kolmogorov-Smirnov test will be performed to determine if the two distributions are equivalent.

An unpaired analysis of all the baseline data from 2020 and the resampled 2021 data will be performed using the following linear model, where 'Cohort' is the year of baseline data:

lm (Y ~ Cohort)

9.3.3 WHO ultrasonography score

An unpaired analysis of all the baseline data from 2020 and the resampled 2021 data will be performed using the following generalised linear model, where 'Cohort' is the year of baseline data:

glm (Y ~ Cohort, family = binomial)

A post hoc decision was taken to not conduct this analysis due to all ultrasound scores being identical in the resampled data and a very small proportion of participants had fibrosis in the 2020 data.

9.4 Demographic and Baseline Variables

The following demographic and baseline variables will be captured in the Demographic CRF, Ultrasonography CRF, Stool sample (1,2 & 3) CRF's & Ultrsonography adjudication CRF and will be summarised according to section 9 using the Full Analysis population.

9.4.1 Demographics

- Age at enrolment: (years)
- Sex: (male/female)
- School: (Buhirigi / Kaiso)
- Class: (class 1, class 2, class 3, class 4)
- Residency in Hoima district :(2 years – 4 years/ > 4 years)

9.4.2 WHO Ultrasonography score

- Niamey Protocol pattern score: (A /B /C / D / E /F)
- Periportal fibrosis: (Yes / No) – derived variable see section 9.2

9.4.3 Schistosoma Infection intensity / infection

Eggs per gram (EPG): – Derived variable see section 9.2

Heavy infection intensity: (Yes / No) – derived variable see section 9.2

Egg Reduction Rate (ERR): – Derived variable see section 9.2

Cure Rate (CR):–Derived variable see section 9.2

Point Prevalence (PP): – Derived variable see section 9.2

9.5 Treatment Compliance

Information relating to the number of participants who missed the IMP at the required visit captured in the 'Non-Adherence to protocol' CRF as well as the 'IMP administration' CRF will be shown in a summary table and displayed as a stacked bar chart, each treatment group represented by their colour and position within the stack. The number missing will be a scale on the y axis and each treatment time-point will be ordered on the x axis.

- Standard of Care Arm: Baseline, 12 months and 24 months
- Intervention Arm 1 (2 x treatment): Baseline, month 7, month 12, month 19 and month 24
- Intervention arm 2 (4 x treatment): Baseline, month 4, month 7, month 10, month 12, month 16, month 19, month 22 and month 24

Information relating to the number of participants who missed data collection visits captured in the 'Non-Adherence to protocol' CRF will be shown in a summary table and displayed using a stacked bar chart with each treatment arm represented by their colour and position within the stack, the number missing will be shown on the y axis and time-point on the x axis.

- Ultrasonography
- Parasitology
    - Standard of Care Arm: Baseline and 24 months
    - Intervention Arm 1: Baseline and 24 months
    - Intervention arm 2: Baseline and 24 months

9.6 Withdrawals

Listings of withdrawals from the trial (ordered by withdrawal reason) will be produced and will consist of the following variables:

- Subject ID
- Treatment group
- Withdrawal reason
- School
- Gender
- Age

- Time to withdrawal (days)

Summary statistics of the:

- withdrawal reason
- school
- gender
- age

will be produced according to section 9 using the safety population.

10 Efficacy Analysis

The primary and secondary efficacy analysis will be analysed using the full analysis population.

The following information will be summarised according to section 9 for the 24 months and 24 months +4 weeks follow up visits using the full analysis population. Box and whisker plots will be used to display continuous endpoints and stacked bar charts to represent binary and categorical endpoints.

- Niamey Protocol pattern score: (A /B /C / D / E //F)*
- Periportal fibrosis: (Yes / No) – derived variable see section 9.2
- Infection intensity: EPG – derived variable see section 9.2
- Infection intensity: Heavy infection – derived variable see section 9.2
- Infection intensity ERR – derived variable see section 9.2
- Infection: Point Prevalence – derived variable see section 9.2
- Cure Rate – derived variable see section 9.2

*Niamey Protocol pattern score will only be presented for the 24 month time point

10.1 Primary Efficacy Analysis

The primary analysis will estimate the absolute difference between each intervention arm and the control arm in the proportion of participants with periportal fibrosis adjusting for the baseline covariates;

- School (Buhirigi / Kaiso)
- fibrosis score – (A, B) /(C ,D, E, F)

- egg count - epg
- duration of residency (2–4 years / > 4 years)

A generalised linear model with a binomial distribution and an identity link function will be fitted (12,13). Standard errors and 95% confidence intervals will be estimated directly from the model.

Should convergence issues arise with the identity link, then a model will be fitted with a logistic link to ensure convergence. The g-formula will then be used to provide an estimate of the average absolute risk difference (12,14). Specifically, for each participant a predicted probability for *all three* arms, but using their original covariate values will be calculated; the within-participant absolute difference calculated; the average of the differences across the participant population will be taken. Bootstrapping will be used to provide a standard error, 95% confidence intervals and p-values.

Hypothesis testing of the null (H0) against the alternative hypothesis (H1) will provide two-sided p-values for each of the three pairwise comparisons between the treatment arms.

H_0: there is no difference in the proportion of participants with periportal fibrosis between treatment arms.

H_1: there is a difference in the proportion of participants with periportal fibrosis between treatment arms.

The multiple testing procedure detailed in section 8.6 will be used whereby the p-value for the first hypothesis is compared to a significance level of 5 % if $p < 0.05$ then the null hypothesis is rejected and the α can be passed to the next hypothesis in the sequence (Fig 2).

A minimum 20% absolute difference in prevalence of periportal fibrosis between the control and intervention arms would be considered a clinically significant outcome. Based on preliminary data an incidence rate of scores C-F of 80% in the control arm is assumed.

10.2 Secondary Efficacy Analyses

10.2.1 Secondary analysis of primary endpoint

A secondary analysis of the primary endpoint will use a GEE (General Estimating Equation) approach (15,16). This will estimate the risk difference in the population adjusting for fibrosis score, epg and clustering by school to take account of any within-school correlation by "fitting' a model which assumes "exchangeable' correlations within schools and will also obtain robust standard errors. The GEE estimand is the risk difference at a population level, whereas it is the risk difference at the individual participant level for the primary analysis.

10.2.2 Analysis of secondary endpoint - Infection intensity (mean epg)

Analysis of the secondary endpoint will test if there is a difference in the mean infection intensity between the control and each intervention arm as well as between each of the two intervention arms.

A multiple linear regression model will estimate the infection intensity at 24 months for each treatment group adjusted for baseline infection intensity (EPG), and school. A z-test will compare each of the three pairwise comparisons separately and provide a p-value for the probability that a difference in the means produced by the model would be observed purely due to chance.

Note that an additional analysis was added post hoc which modelled the mean infection intensity by treatment group which adjusted for baseline infection intensity and school with an interaction term between treatment group and school.

10.2.3 Analysis of secondary endpoint - Infection intensity (Heavy infection)

An additional analysis of the infection intensity endpoint will be conducted by dichotomising EPG into ≥400 EPG or <400 EPG relating to 'heavy infection' and 'not heavy infection'. The analysis will estimate the proportion of 'heavy infection' for each treatment arm by fitting a logistic regression model and adjusting for baseline infection intensity (epg), and school. 95% confidence intervals for the proportions will be estimated using the standard error from the model.

10.2.4 Analysis of secondary endpoint - Egg Reduction Rate

A mixed model repeated measures analysis with an unstructured covariance matrix for the within - patient residual errors will estimate the Egg Reduction Rate (ERR) between:

- baseline and 24 month +4 weeks
- baseline and baseline + 4 weeks
- 24 months and 24 months + 4 weeks

for each treatment group using log transformed epg counts at 3 time points (baseline +4 weeks, 24 months and 24 months + 4 weeks). The model will specify fixed effects of treatment, time and baseline epg count with an interaction term between treatment and time. A weight argument will allow a distinct variance for each time point. The treatment effect coefficient (βx) in this model quantifies the relative change and the ERR is given by 1- exp (βx).

10.2.5 Analysis of secondary endpoint - Prevalence of infection

The burden of disease will be estimated by calculating the prevalence at 24 months. A binary variable indicating disease status will be created and each participant will be classed as either having Schistosomaisis or not using the average EPG at 24 months.

No Disease 0 = EPG = 0

Disease 1 = EPG > 0

A logistic regression model will be fitted to estimate the prevalence odds ratio adjusted for age, sex and school.

Note that an additional analysis was included post hoc which modelled prevalence of infection at 24 months adjusting for baseline ln(epg+1), age, sex and school.

10.2.6 Analysis of secondary endpoint - Cure rate

The cure rate as defined in section 9.2 will be calculated for the following time points:

- Baseline to 24 months
- Baseline to baseline +4 weeks
- 24 months to 24 months + 4 weeks

Cure rates will be presented by treatment arm and also stratified by baseline infection intensity; ≥400 epg = Heavy infection / <400 – 1 epg = Medium – Light Infection.

10.2.7 Alluvial plots

Alluvial plots will be produced to show changes in the proportions of participants who have epg >0 and epg = 0 over time for:

- All participants combined
- Each treatment group separately

Alluvial plots will also be produced to show changes in the proportions of participants who have Heavy infection (epg >400) over time.

- All participants combined
- Each treatment group separately

10.3 Exploratory Efficacy Analyses

10.3.1 Complier Average Causal Effect analysis

To investigate how treatment compliance affects the primary outcome CACE (Complier Average Causal Effect) analysis will be performed (17) to compare each intervention arm separately to the control arm.

The CASE will be calculated using the ITT treatment effect and the proportion of compliers in the treatment arms, different levels of compliance will be used and are defined below for each intervention arm:

Intervention arm 1: Participants who take ≥ 4 doses of the PZQ out of a maximum of 5 possible doses.

Intervention arm 2: Participants who take ≥ 8 doses of the PZQ out of a maximum of 9 possible doses.

AND

Intervention arm 1: Participants who take > the baseline i.e. the control arm dose of PZQ (2 doses)

Intervention arm 2: Participants who take > intervention arm 1 dose of PZQ i.e. 5 doses

CASE calculation:

$$\frac{\text{Intention To Treat (ITT) treatment effect}}{\text{proportion of compliers in the intervention arm}}$$

The CACE estimation relies on the following assumptions:

- Treatment assignment is random
- Merely being offered the treatment has no effect on the outcomes
- If participants in the control arm had been randomised to the intervention arm, they would comply with the intervention at the same rate as those actually randomised to the intervention
- There is an absence of defiers – patients that take the opposite treatment to what is offered
- Stable Unit Treatment Value Assumption (SUTVA). SUTVA requires that the response of a particular participant depends only on the treatment to which the participant was assigned, not the treatments of others.

If these assumptions hold we can also estimate the potential outcomes among participants in the control arm using simple algebra.
'Complier' and 'Non-complier refer specifically to compliance to the intervention and participants in the control arm are assumed to receive the control treatment as intended and values therein are based on the aforementioned assumptions.

This analysis will also be performed for the secondary outcome of infection intensity:

- mean epg
- Heavy infection

11 Safety Analyses

The safety population will be used for all safety analysis.

Safety data collected on the day of treatment and at 24 hours will be grouped for each time point. i.e. AE's from day 1 and +24 hours will be grouped for baseline AE's and any repetitions will not be counted within the time-point.

11.1 Adverse Events

11.1.1 Adverse Events listings

A table of Adverse Event listings will be produced with the following headings:

- **Participant ID:**
- **Randomised assignment:**
- **Adverse Event Number:**
- **AE term:** 1=Headache 2=Vertigo 3=Seizures 4=Abdominal pains with nausea 5=Abdominal pains without nausea 6=Nausea without abdominal pains 7=Diarrhoea 8=Rash 9=Pruritus 10=Facial oedema 11=Urticaria 12=Vomiting 13=Other
- **Date of Onset:** DD/MM/YYYY
- **Date of resolution:** DD/MM/YYYY
- **Outcome:** 1=Resolved without Sequelae 2=Resolved with Sequelae 3=Ongoing 4=Death 5=Unknown
- **Severity:**
- **Seriousness:** 0=Not Serious 1=Resulted in death 2=Life-threatening 3=Required inpatient or prolonged existing hospitalisation 4=Resulted in persistent or significant disability/incapacity 5=Resulted in congenital anomaly/birth defect 6=Other Important Medical Event, please specify
- **Causality:** 1=Definitely related 2=Probably related 3=Possibly related 4=Unlikely to be related 5=Not Related
- **Treatment:** Y/N

Listings will be ordered by subject ID and timepoint.

11.1.2 Adverse Events summary statistics

Summary statistics of the frequency and percentage of AE's will be presented by treatment arm and time point for all treatment time points. Note that not all treatment arms treated at every time point thus Adverse Events information will not be reported unless treatment has occurred in error. Summary statistics will be provided for;

- all adverse events grouped
- for each Averse Event term (1–13) separately
- AE's grouped by system i.e. Neurological and Digestive

Comparisons of the frequency of adverse events between treatment arms can be made at baseline, 12 months and 24 months. The relative risk comparing the treatment arms separately to the control arm and comparing the two treatment arms to each other for an AE grouping will be presented if the number of adverse events is > 10% for any treatment arm.

Dot plots will be used to show the frequency, percentage and relative risk for each of the 3 treatment arm comparisons at each time point.

11.1.3 Analysis of adverse events and infection intensity

11.1.3.1 Exploratory analysis of Adverse Events and infection intensity: ROC curves

This analysis will produce Receiver Operator Characteristics (ROC) curves (18) to graphically represent the connection between clinical sensitivity and specificity for every possible cut off of infection intensity (epg) as a 'test' for predicting Adverse Events .

Adverse Events will be considered as follows:

1. Any adverse event (1–13) experienced, as either having an AE or not (Y/N)

2. Any digestive adverse event (4, 5, 6, 7 or 12) as either having an AE or not (Y/N)
3. Any neurological adverse event (1, 2, 3, 8, 9 or 11) as either having an AE or not (Y/N)

For each ROC curve the Area under the ROC will be calculated to provide an estimate of the efficiency of using infection intensity (epg) in predicting AE occurrence.

ROC curves will be produced for baseline and 24 month time points.

11.1.3.2 Exploratory analysis of Adverse Events and infection intensity: Logistic regression

A further exploratory analysis will model the log odds of any adverse event using infection intensity (epg) on a continuous scale, treatment, and age group as

predictors. Exposure variables will be modelled as fixed effects with an interaction term between infection intensity and age group.

Logistic regressions for both baseline and 24 month time points will be performed.

11.2 Deaths, Serious Adverse Events and other Significant Adverse Events

Any AE given a seriousness score of 1- 6 will be considered an SAE and any SAE having a definite, probable or possible causality will be considered an SAR.

The number of patients per arm with SAE/Rs will be counted and the frequency as well as the percentage will be presented at baseline and 24 months. The relative risk and 95% CI's of the 2x and 4 x treatment arms compared with the standard treatment will also be presented if > than 10% events are observed in any of the treatment arms.

12 Figures

- CONSORT diagram showing patient disposition
- Stacked bar chart showing treatment compliance
- Stacked bar chart showing assessment compliance
- A graph showing the proportion of missing data for key baseline and outcome variables
- Bland-Altman plot comparing epg for split baseline March 2020 and March 2021
- Bar chart showing proportion of periportal fibrosis by treatment group
- Stacked bar chart showing proportion of each Niamey Protocol pattern score by treatment group
- Box and whisker plots showing the log transformed infection intensity EPG by treatment group at baseline and 24 months
- Bar chart showing the proportion of 'heavy infection intensity' by treatment group
- Box and whisker plots of ERR by treatment group for each time-point
- Alluvial plots showing epg status (epg= 0 / epg >0) over timepoints
- Dot plot to show the frequency of AE's comparing each pair of treatment arms at baseline and 12 months.

- Receiver Operator Characteristics (ROC) curves for adverse events at baseline and 24 months

13 Reporting Conventions

P-values ≥0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". The mean, standard deviation, and any other statistics other than quantiles, will be reported to one decimal place greater than the original data. Quantiles, such as median, or minimum and maximum will use the same number of decimal places as the original data. Estimated parameters, not on the same scale as raw observations (e.g. regression coefficients) will be reported to 3 significant figures.

14 Technical Details

R version 4.1.1 will be the software tool used, correct at the time of writing. Full documentation of all versions and add-on packages will be recorded as the report is generated.

Any outputs will have:

- The date and time included
- The name of the code file that produced the analysis
- The author
- A log capturing the version of the software and any external add-on code used.

15 Summary of Changes to the Protocol

15.1.1 Protocol v 2.0

- Updated CI Maternity Cover (Front page, Section 2 Trial Management Committees and Section 26.3 Safety Reporting Flow Chart)

- Addition of 'school-community based' to aim (Section 7) to enable trial activities to be carried out whilst schools are closed due to COVID-19 National Guidelines

- Addition Section 9 Community School and Trial Team COVID-19 precautions.

- Section 11.1.1.2 removal of IMP expiry date of 31/01/2022, as trial will run beyond this date due to COVID-19 pause of trial activities.

- Addition of 'local community trial site location' and 'house-hold to house-hold' throughout the protocol to enable trial activities to be carried out in the appropriate locations whilst the schools are not permitted to be used for trial activities due to COVID-19 restrictions. (Section 12)

- Explicit statement that trial activities to return to the participating school location when allowed. (Section 12)

- Clarification of exploratory outcome sample collection (Section 12.3.3).

- Addition of participant photograph to enhance participant identification for community based trial activities (Section 12.3.3.1)

- Widening of window between baseline assessments and IMP administration so that baseline data may be used. Delay between activities caused by COVID-19 pause to trial activities. (Section 12.4)

- Addition Section 12.4.1 Baseline Assessment Assurance following COVID-19 Pause. Text added to allow repeat collection of primary and secondary endpoint data from Buhirigi School participants to verify data integrity despite time delay between first collections and IMP administration caused by COVID-19 pause.

- Section 12.5.3 Addition of footnotes to the Schedule of assessments to allow for wider window between ultrasound examination and first IMP administration at baseline visit, due to COVID-19 pause.

- Widening of temperature storage range for collected samples (Section 16)

15.1.2 Protocol v 3.0

Schedule of assessments and Trial assessments following the 1st treatment dose

- Update of 1-year procedures in trial synopsis (section 4) to reflect removal of ultrasound examination and reduction in sampling for egg counts.
- Update of trial flow chart (section 5) to reflect removal of 1-year+4-week egg count for exploratory outcome of egg reduction rate

- Update of section 8.6. Trial outcome measures to reflect altered examinations/sampling schedule detailed above for changes to sections 4 and 5, plus removal of blood sampling at 1-year.
- Update of Schedule of assessment (Section 12.6) to reflect altered examination/sampling schedule detailed above for changes to sections 4 and 5.
- Update of Response Criteria (Section 15.1) to reflect removal of ultrasound examinations at 1-year.

Update of appendix 1 to reflect changes in Data Management responsibilities

15.1.3 Protocol v4.0

Sumbitted to VCD REC prior to NDA approval

- Update of Trial Management Committee (section 2) to reflect change in Trial Management Group
- Update of Trial Synopsis (section 4) with clarification of AE monitoring procedure
- Update of Reporting Serious Advers events (section 13.5) to reflect new SAE reporting structure
- Update of Toxicity (section 14) to reflect change in qualification required for the Trial Site Manager
- Update of Participant registration/ Randomisation procedure (section 26.1.1) to correctly reflect procedure that took place

Update of Safety Reporting Flow Chart (Section 26.4) to reflect changed qualification requirement of Trial Site Manager and where the medical oversight is provided.
Sumbitted to VCD REC prior to NDA approval

15.1.4 Protocol v4.1

Administrative change, submitted to VCD REC prior to NDA approval (note – submitted prior to review of version 4)

- Update of Trial management committee (section 4) to include name of Trial Site manager

Update of Safety Reporting Flow Chart (Section 26.4) to include name of Trial Site manager.

16 Summary of Changes to the SAP

16.1.1 Re-estimation of the power of the study based on a revised prediction of the number of participants remaining in the trail by 2 years

It was felt that the COVID-19 pandemic and missing consent documentation leading to the necessity of re-consenting around 25% of individuals would lead to a much reduced sample size. Therefor a different approach to multiple testing of the primary analysis hypotheses was taken which allowed a 5% alpha to be used.

A revised 100 participants per treatment group was used to estimate the power when the treatment effect was a 20%, 25% and 30% reduction in the primary outcome.

16.1.2 Change to definition of Full Analysis population

Removal of the condition to only include participants 'treated at baseline' in the population. Due to the COVID 19 pandemic many participants were not treated at baseline but were treated at other time points in the trial. Including only participants treated at baseline would significantly reduce the sample size and reduce the power. Also information would be lost for those participants not treated at baseline but treated at other time points

16.1.3 Creation of category A/B for WHO Ultrasonography score

For participants who had real and recorded periportal fibrosis scores recorded as A and B or B and A these participants will be categorised as having a score of A/B in any table reporting on the summary statistics for the WHO Ultrasonography score.

16.1.4 Dealing with missing baseline data for WHO Ultrasonography score and egg count data

There is a proportion of baseline data missing from the ultrasound and egg counts due to the delayed start of the trial due to the COVID-19 pandemic. Removal of these participants from the data set would potentially lead to removal of outcome data from the analysis models and less efficiency overall. We will impute values for these missing values as described in section 8.4

16.1.5 Post Hoc analysis added

a. Multiple linear regression of infection intensity at 24 months on treatment group: This analysis was repeated with an interaction term between school and treatment group.
b. Logistic regression of the prevalence of infection at 24 months on treatment group. This analysis was repeated to include baseline egg count as a covariate.

16.1.6 Baseline Resample comparison – Ultrasonography

This analysis was not performed since all resample data was identical.

16.1.7 Sensitivity Analysis

6 participants who required adjudication of their baseline ultrasonography scores did not receive an adjudicated score as the images required to perform this process could not be located. It was therefor decided that a best case/worst case approach to a sensitivity analysis be performed. A total of 6 additional data sets were created where different possible combinations of the missing baseline scores were added. 1 of the participants only had one image score and this was used as the adjudicated scores in all data sets. 2 participants did not have final ultrasound data at 24 months and thus were not included in any analysis.

17 References

1. Cheever AW. A Quantitative Post-Mortem Study of Schistosomiasis Mansoni in Man. Am J Trop Med Hyg. 1968;17(1):38-64.

2. Organization WH, Crompton DWT. Sustaining the Drive to Overcome the Global Impact of Neglected Tropical Diseases : Second WHO Report on Neglected Tropical Diseases. Geneva, SWITZERLAND: World Health Organization; 2013.

3. Zhang Y, Koukounari A, Kabatereine N, Fleming F, Kazibwe F, Tukahebwa E, et al. schistosomiasis and soil-transmitted helminthiasis in Uganda. 2007;11:1-11.

4. Ongom VL, Bradley DJ. The epidemiology and consequences of Schistosoma mansoni infection in West Nile, Uganda: I. Field studies of a community at Panyagoro. Trans R Soc Trop Med Hyg. 1972;66(6):835-51.

5. Booth M, Vennervald BJ, Kabatereine NB, Kazibwe F, Ouma JH, Kariuki CH, et al. Hepatosplenic morbidity in two neighbouring communities in Uganda with high levels of Schistosoma mansoni infection but very different durations of residence. Trans R Soc Trop Med Hyg. 2004;98(2):125-36.

6. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear technique in Schistosomiasis mansoni. Rev Inst Med Trop Sao Paulo. 1972;14(6):397-400.

7. Berhe N, Myrvang B, Gundersen SG. Reversibility of Schistosomal Periportal Thickening/Fibrosis after Praziquantel Therapy: A Twenty-Six Month Follow-up Study in Ethiopia. The American Journal of Tropical Medicine and Hygiene Am J Trop Med Hyg. 78(2):228-34.

8. Carpenter JR, Kenward MG. Missing data in randomised controlled trials: a practical guide. Health Technology Assessment Methodology Programme; 2007.

9. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med. 2005;24(7):993-1007.

10. Stef van Buuren. Flexible Imputation of Missing Data. Chapman and Hill. 2018.

11. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586–604.

12. Naimi AI, Whitcomb BW. Estimating Risk Ratios and Risk Differences Using Regression. Am J Epidemiol. 2020;189(6):508–10.

13. Richardson TS, Robins JM, Wang L. On Modeling and Estimation for the Relative Risk and Risk Difference. J Am Stat Assoc. 2017;112(519):1121–30.

14. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. Am J Epidemiol. 2011;173(7):731–8.

15. Hanley JA, Negassa A, Edwardes MD deB, Forrester JE. Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. Am J Epidemiol. 2003;157(4):364–75.

16. Hubbard AE, Ahern J, Fleischer NL, Laan M Van Der, Satariano SA, Jewell N, et al. the Associations Between Neighborhood Risk Factors and Health Linked references are available on JSTOR for this article : _ Original Article _ To GEE or Not to GEE Comparing Population Average and Mixed Models for Estimating the Associations Between Neigh. 2010;21(4):467–74.

17. Simmons RA, Egger JR. A CACE for assessing non-compliance: Addressing non-compliance in randomised controlled trials. BJOG. 2019;126(13):1558–9.

18. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022;75(1):25–36.

18 Listing of Tables, Listings and Figures

meta_table v0.1.xlsx