A randomized, placebo-controlled, double-blinded trial of the safety and efficacy of tecovirimat for the treatment of adult and pediatric patients with monkeypox virus disease

Statistical Analysis Plan

Version 2.0

21 February 2024

NCT05559099

STATISTICAL ANALYSIS PLAN
for
PALM007

Study Title:

A randomized, placebo-controlled, double-blinded trial of the safety and efficacy of tecovirimat for the treatment of adult and pediatric patients with mpox virus disease (PALM 007)

Version 2.0

DATE: 21 February 2024

Study Type:	Randomized, placebo-controlled, double-blinded clinical trial
Products:	tecovirimat (TPOXX®)
Drug Sponsor:	SIGA Technologies, Inc.
Study Sponsors:	National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), United States of America, and Institut National de la Recherche Biomédicale (INRB), Democratic Republic of the Congo (DRC)

SIGNATURE PAGE

SPONSORS: National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), United States of America, and Institut National de la Recherche Biomédicale (INRB), Democratic Republic of the Congo (DRC)

STUDY TITLE: A randomized, placebo-controlled, double-blinded trial of the safety and efficacy of tecovirimat for the treatment of adult and pediatric patients with mpox virus disease (PALM 007)

SIGNATURES

Individual / Role	Affiliation	Signature	Date
Placide Mbala, MD, PhD **DRC Co-principal Investigator**	INRB		
Olivier Tshiani Mbaya, MD, MTM&GH **US Protocol Co-chair**	Clinical Monitoring Research Program Directorate (CMRPD) Frederick National Laboratory Leidos Biomedical Research, Inc.		
Veronique Nussenblatt, MD, ScM, M.H.S **US Protocol Co-chair**	NIAID		
Tyler Bonnett, MS **Statistical Lead (Blinded)** **SAP Author**	CMRPD Frederick National Laboratory Leidos Biomedical Research, Inc.		
Lori Dodd, PhD **Statistical Co-Lead (Blinded)**	NIAID		

STATISTICAL ANALYSIS PLAN REVISION HISTORY

Version Number	Version Date	Corresponding protocol version and date	Summary of Changes with Rationale
1.0	22 August 2023	Version 3.0; 16 August 2023	Initial Version
2.0	21 February 2024	Version 4.0; 15 December 2023	• Added Appendix 2 detailing updates to the planned sample size and interim monitoring plans. Existing sections detailing the original plan were left as-is but a note was added pointing the reader to Appendix 2 for the most recent plan (Sections 2.1, 5.3, and 8.2) • Updated Section 6.4 Primary Endpoint Analysis Details clarifying rules for determinination of the primary endpoint in cases where participants develop new lesions in the study period after already meeting the protocol definition of lesion resolution. • Updated Sections 6.5.2 and 6.6.6 describing planned analyses of PCR data to clarify how analyses will handle cases where the clade I result is negative but the generic orthopox result is positive.

Note: Protocol updates may not necessarily require SAP revisions. Each version of the SAP will note the protocol version in use on the date the SAP was revised.

Table of Contents

1 Preface

This Statistical Analysis Plan (SAP) for "A randomized, placebo-controlled, double-blinded trial of the safety and efficacy of tecovirimat for the treatment of adult and pediatric patients with mpox virus disease" (PALM 007) describes and expands upon the statistical information presented in the protocol. This document describes all planned analyses and provides reasons and justifications for these analyses.

2 Study Background

Mpox virus (MPXV) belongs to the genus *Orthopoxvirus* (OPXV) and is antigenically related to *Variola*, the causative agent of smallpox, and *Vaccinia* viruses. There are two *MPXV* clades, each one associated with different clinical expressions and geographic locations. Clade 1 (formerly called the Congo Basin clade) has an estimated case-fatality rate of 11%. Clade 2 (formerly called the West African clade) is reported to be less severe and has an estimated case-fatality rate of 0-6%. While cases outside of West and Central Africa were previously rare, there is an ongoing (as of August 2022) worldwide outbreak of mpox linked to the clade 2 MPXV which has resulted in thousands of cases, predominantly in Europe. The DRC (where clade 1 is more prevalent) has reported more than 1000 cases of mpox each year since 2005 and cases have been increasing in sub-Saharan Africa since 2000 [1]. The similarity between MPXV and the variola virus, coupled with the high comorbidity on affected individuals from areas with limited resources, have placed mpox treatments at the forefront of public health and scientific research agendas in many countries.

The clinical features of the disease are characterized by two phases: the prodromal and rash phases. The prodromal phase includes general signs like fever, headache, chills, sweats, sore throat, myalgias, prostration, and lymphadenopathy. The rash phase can last between 2 and 4 weeks. Skin lesions are painful and progress uniformly from macules to papules, vesicles, and pustules. This process eventually results in umbilication, scabbing, and finally desquamation. Lesions have a centrifugal distribution starting from the head and the face of the patient with extension to the trunk and the extremities of the body.

To date, most of the patients in remote regions affected by mpox receive only supportive and symptomatic care as standard treatment. However, several investigational antivirals, initially developed against smallpox, demonstrate activity against MPXV and other *OPXV* in vitro and in animal models [2], but none has been evaluated in a clinical trial. The most promising is the antiviral tecovirimat (TPOXX®) developed by SIGA Technologies, Inc. (SIGA) and approved by the US Food and Drug Administration (FDA) for the treatment of smallpox [3,4]. Tecovirimat is also authorized by the European Medicines Agency for the treatment of smallpox, mpox, and cowpox.[22]

This is a randomized, placebo-controlled, double-blind study to test the safety and efficacy of tecovirimat in adults and children with laboratory-confirmed mpox virus (MPXV) disease at study sites in the DRC. Participants will be randomly assigned to receive oral tecovirimat or placebo (1:1 via block randomization, stratified by study site and days from onset of prodromal symptoms ≤7 days or >7 days), each administered in the hospital with standard-of-care (SOC) treatment for 14 days. Participants will be followed for 28 days with an optional visit at Day 59 for long-term assessment.

2.1 Purpose and timing of the analyses

NOTE: The text below describes the original sample size calculations and interim monitoring plan. See Appendix 2 for details on the updated sample size and timing of planned analyses.

This SAP encompasses all interim analyses and the final analysis of primary, secondary, and exploratory outcome measures. These analyses will assess the efficacy and safety of tecovirimat compared to placebo. The primary objective of the study is to evaluate and compare time to lesion resolution, defined as the first day on which all lesions on the body are scabbed or desquamated or a new layer of epidermis has formed. The number of lesion resolution events observed is a critical marker of trial progress. As detailed in Section 5.3, 318 events are required to adequately power the trial assuming the hypothesized rate ratio for lesion resolution of 1.40 (that is, a 40% better rate of resolution on the tecovirimat arm relative to the placebo arm). The protocol specifies that planned interim efficacy and futility analyses will be performed after 1/3 and 2/3 of the targeted number of events required. Thus, the planned interim analyses will occur after 106 and 212 lesion resolution events have been observed. The goal of the interim analyses is for the DSMB to review safety and efficacy data and provide a recommendation on whether the study should proceed or stop early for either evidence of benefit or futility. A final analysis will be performed at the end of the study and will report the results of all prespecified analyses contained in this SAP.

3 Study Objectives

3.1 Primary Objective

To evaluate the clinical efficacy, as assessed by time to lesion resolution, of tecovirimat plus SOC versus placebo plus SOC for patients with mpox.

3.2 Secondary Objectives

1. To evaluate the clinical efficacy, as assessed by time to lesion resolution, of tecovirimat plus SOC versus placebo plus SOC for patients with mpox, according to duration of symptoms (≤7 days or >7 days).
2. To evaluate the virologic efficacy, as assessed by PCR separately of blood, skin lesion, and oropharynx samples, of tecovirimat plus SOC relative to placebo plus SOC for patients with mpox.
3. To evaluate the clinical efficacy of tecovirimat plus SOC versus placebo plus SOC in patients with mpox as assessed by mortality, clinical severity, and duration of symptoms.
4. To evaluate the safety of tecovirimat plus SOC relative to placebo plus SOC in patients with mpox.

3.3 Exploratory Objectives

1. To evaluate the frequency and characteristics of persistent residual lesions.
2. To describe lesion progression longitudinally over the study period.
3. To evaluate exposure history of confirmed mpox cases and to identify risk factors for MPXV infection.

4. To develop a baseline disease severity metric for mpox.
5. To evaluate the potential impact of the presence of anti-OPXV antibodies on the course of disease and the clinical efficacy of tecovirimat.
6. To evaluate persistence of MPXV PCR positivity in blood, skin lesions, and the oropharynx.
7. To evaluate the trajectory of MPXV IgM and IgG over time during infection.
8. To assess genomic variability in MPXV isolated from participants based on geographic and clinical differences.
9. To assess whether viral resistance develops due to selective pressure by treatment.
10. To assess the effect of HIV infection on mpox clinical outcomes and treatment effect.
11. To determine tecovirimat drug levels in a real-world scenario.
12. To describe clinically and virologically any cases of recrudescent disease as defined by the protocol

4 Study Endpoints

4.1 Primary Endpoint

Time to lesion resolution, defined as the first day on which all lesions on the total body are scabbed or desquamated or a new layer of epidermic has formed, up to 28 days after randomization.

4.2 Secondary Endpoints

1. Time to lesion resolution, as defined in the primary endpoint, according to stratification by time from onset of illness as ≤ 7 days or > 7 days.
2. Proportions with negative PCR results separately by blood, oropharyngeal swab, and lesion swab samples 14 days after randomization.
3. Mortality within the first 28 days after randomization.
4. Time to death up to 28 days after randomization.
5. Frequency and duration of clinical symptoms (including nausea, vomiting, abdominal pain, diarrhea, anorexia, cough, lymphadenopathy, dysphagia, sore throat, muscle aches, fatigue/lack of energy, fever, chills, night sweats, headache, ocular lesions, eye pain, change in vision, buccal ulcers, nasal congestion, cough, joint pain, pain with urination, painful skin lesions, pruritic skin lesions).
6. Incidence of SAEs, AEs requiring drug discontinuation, and incidence of other AEs.
7. Incidence of bacterial infections. Bacterial infections will be defined clinically with laboratory and radiographical confirmation when possible.
8. Automated image analysis of lesion counts and characteristics over time.

4.3 Exploratory Endpoints

1. Presence, location, and duration of persistent residual lesions, defined as any lesion (in any area of the body) unresolved after all assessment-region lesions are scabbed or desquamated.
2. Longitudinal description of lesion progression over the study period.
3. Number and percentage of confirmed mpox cases reporting exposure to animals, symptomatic humans, or with no known exposures.
4. Associations between measures of baseline disease severity (including lesion counts, duration of symptoms, and comorbidities) and efficacy endpoints.
5. Differences in outcomes based on the presence of anti-OPXV antibodies.
6. Viral load and proportion with negative PCR results in blood, oropharyngeal swab, and open lesion swab samples over time and time to the first negative PCR result in each of these specimens up to 28 days after randomization.
7. Change in antibody titer over time.
8. Sequencing differences between virus isolated from participants from different geographic areas and with differing clinical trajectories and responses to tecovirimat.
9. Sequencing differences between virus isolated prior to and after treatment.
10. Differences in severity and duration of MPXV infection and treatment effect by HIV status, including HIV viral load.
11. Pre-dose concentration of tecovirimat in blood measured on day 7.
12. Incidence of recrudescent disease and description of clinical and virologic characteristics of recrudescent disease cases.

5 General Statistical Considerations

5.1 Overview

This is a randomized, placebo-controlled, double-blind trial with a two-sided type I error rate of 5% for the primary endpoint.

To date there have been no studies of mpox therapeutics in humans to inform the design of the present trial. As such, there are no standards for study endpoints. The primary endpoint was largely informed by subject matter experts with experience treating mpox and secondary analysis of data generated from an observational study of mpox patients in DRC carried out by the DRC's INRB in collaboration with the USAMRIID between 2007 and 2011 [5].

Prespecified analyses assume the availability of data. However, given the rural location of study sites, it is possible that field circumstances may arise that inhibit data collection (e.g., equipment for malaria and HIV tests may not be available on site and tests may have to be completed retrospectively). Prespecified analyses that are not possible due to data limitations may be omitted or adapted. Any deviations from the prespecified analyses will be documented.

5.2 Primary Endpoint Selection

The primary endpoint emphasizes the clinical relevance of lesion resolution for patients with mpox. Mpox lesions are painful and reduce patients' ability to carry out ordinary tasks such as swallowing. Lesions progress through discrete stages at roughly the same rate starting with macules (flat lesions) and moving on to papules (raised), vesicles (raised and filled with clear fluid), pustules (filled with opaque fluid), and finally scab and fall off (desquamate). Lesion scabbing and desquamation generally coincides with resolution of other symptoms and is an important criterion for discharge in clinical practice (in addition to blood PCR negativity, which is included as a secondary endpoint). Therefore, demonstration of an improvement in time to lesion resolution would provide a direct clinical benefit to patients. The analysis of the primary endpoint will consider the competing risk of death. The key measure is the subdistribution rate ratio (also called a subdistribution hazard ratio; these terms are used synonymously throughout this document)—a measure of the instantaneous rate of lesion resolution for individuals who have not yet reached resolution either because they are alive with unresolved lesion or because they died prior to achieving lesion resolution. A 40% improvement in the rate of lesion resolution (corresponding to a subdistribution rate ratio of 1.40 comparing tecovirimat to placebo) is considered the minimal clinically meaningful difference—smaller differences are not thought to be of interest.

5.3 Power and Sample Size

NOTE: The text below describes the original sample size calculations. See Appendix 2 for details on the updated sample size.

For the test of the primary endpoint, the two key determinants of power are the total number of lesion resolution events, *E*, and the treatment-to-control ratio of the rate of lesion resolution, θ. The number of events required to achieve power $1-\beta$ to detect a subdistribution rate ratio of θ using a two-tailed test with type I error rate α=0.05 is

$$E = \frac{4\left(z_{\alpha/2} + z_{\beta}\right)^2}{\{\ln(\theta)\}^2},$$

where z_x is the $100(1-x)$th percentile of the standard normal distribution [6,7].

Table 1 displays the power of this test for various scenarios. In total, 318 participants with lesion resolution up to 28 days after randomization are needed to detect a 40% improvement in the rate of lesion resolution as measured by the subdistribution rate ratio with 85% power and a two-sided type one error rate of α=0.05. Data from a previous observational study of mpox in the DRC revealed a 22-day lesion resolution event rate of 77%. Extrapolating that rate yields a total targeted sample size of 413 participants to achieve 318 events. Although every effort will be made to eliminate patient dropout, a total sample size of 450 is planned to account for potential dropout.

Note that the event rate used to determine the sample size is derived from an observational cohort and thus is most likely to correspond to the event rate on the control arm in this trial. Assuming the hypothesized sub-distribution rate ratio of 1.40 is accurate, the event rate on the tecovirimat arm (and therefore the overall event rate) will be higher than the rate used to estimate the sample size. However, because the trial is event-driven, the somewhat conservative approach to estimating the target sample size is not a critical issue—the trial will enroll participants until 318 lesion resolution events have occurred.

Table 1. Number of events needed for 80% and 85% power. The shaded region represents the rate ratio, number of events, and projected sample size for the trial before accounting for potential drop out.

Subdistribution Rate Ratio (θ)	Scenario for 80% Power (β=0.20)		Scenario for 85% Power (β=0.15)	
	Number of Events Needed	Number of Participants Needed*	Number of Events Needed	Number of Participants Needed*
1.15	1608	2089	1839	2389
1.20	945	1228	1081	1404
1.25	631	820	722	938
1.30	457	594	522	678
1.35	349	454	399	519
1.40	278	362	318	**413**
1.45	228	297	261	339
1.50	191	249	219	285
1.55	164	213	187	243
1.60	143	186	163	212
1.65	126	164	144	188

* Assumes that 77% of participants will experience the event and does not account for potential dropout.

5.4 Sample Size Re-estimation Procedure

A blinded sample size re-estimation will be conducted at the midpoint of the trial when 159 lesion resolution events have occurred (i.e., at 50% information). The overall rate of events combining both study arms will be computed and used to determine whether the planned sample size should be increased either to achieve the 318 total events needed to power the trial under the original design or to increase overall power (i.e. to target greater than 318 events, which may be desirable if enrollment allows). The denominator for calculating the event rate should consider the fact that recently enrolled participants may not have had time to experience events. One reasonable denominator of choice would be the number of participants with potential for at least 28 days of follow-up (i.e., the full planned. observation period). If lesion resolution events tend to occur early, for example if the large majority of events occur within 14 days, then including participants with 14 or more days of potential follow-up would provide a reasonable estimate. Another approach would be to use the Kaplan-Meier based estimate of the 28-day rate. The blinded sample size re-estimation committee referenced below may request multiple estimates of the event rate used to potentially increase the sample size.

For example, if at the time 159 lesion resolution events have been observed there have been 265 participants with potential for 14 days of follow-up, the overall event rate is 159/265 = 60%. In this case the suggested number of participants would be 318/0.60 = 530 rather than the 450 planned at the start of the trial. The sample size increase computation in this scenario is a suggestion rather than a strict rule. The sample size re-estimation committee may, for instance, recommend increasing the sample size by slightly more than the suggested amount.

As mentioned above, the committee may also choose to increase the target number of events and therefore the overall power of the study. For example, such a decision may be appropriate if

enrollment is steady since the operational difficultly posed by the study setting argues in favor of fully utilizing the trial infrastructure and incorporating data from as many participants as possible aligns with the goal of providing definitive evidence. Increasing the target number of events for the trial would have an impact on interim monitoring bounds. The decision to increase the sample size would be accompanied by a plan for handling downstream effects on future analyses and the overall type I error rate. Monitoring boundaries will be computed using the Lan-DeMets alpha spending function analog of the O-Brien-Fleming boundaries (see Section 8.2), which is a flexible procedure that can accommodate the changes to interim analysis timing that a sample size increase could necessitate.

5.5 Randomization

Randomization will be performed onsite by an unblinded pharmacist. Participants will be randomized via permuted block randomization in a 1:1 ratio to tecovirimat or placebo within randomization strata of days from onset of prodromal symptoms (≤7 days or >7 days) and study site. The unblinded statisticians will prepare an allocation plan detailing the block sizes and probabilities used to create the final allocation table. The code for creating the allocation table will be stored securely. No details of the allocation plan or code will be shared outside of the unblinded statistical team until the study is complete.

A secure online portal will be used for all randomizations if possible. The unblinded pharmacist will access this system and enter the participant information (including the date of onset of prodromal symptoms, as required for determining the correct stratum). The system output is a masked treatment code, which is then used to determine the participant's study assignment. The unblinded pharmacist will be provided with a mapping of masked treatment codes to actual study product names prior to study start.

Because of the rural location of study sites, a consistent and strong internet connection is not guaranteed. While randomization using the online tool is preferred, a backup procedure for randomization via sequentially numbered secure envelope will be in place at study sites. Backup randomization envelopes for each stratum will be shipped to the site in advance of study start in secure boxes and stored in an access-controlled location. Because there is no way to predict when (if at all) these randomization envelopes will be needed, the assignments provided by the randomization envelopes will come from a separate allocation table from the table used in the online system. Both allocation tables will use 1:1 permuted block randomization, but the block sizes and probabilities may differ. The allocation plan will contain these details.

5.6 Hierarchical Testing Scheme

The study will employ a hierarchical testing scheme for the primary endpoint (time to lesion resolution) in order to maintain an appropriate type I error rate while still allowing for the possibility of proceeding to within-stratum tests of time to lesion resolution as part of the first secondary endpoint.

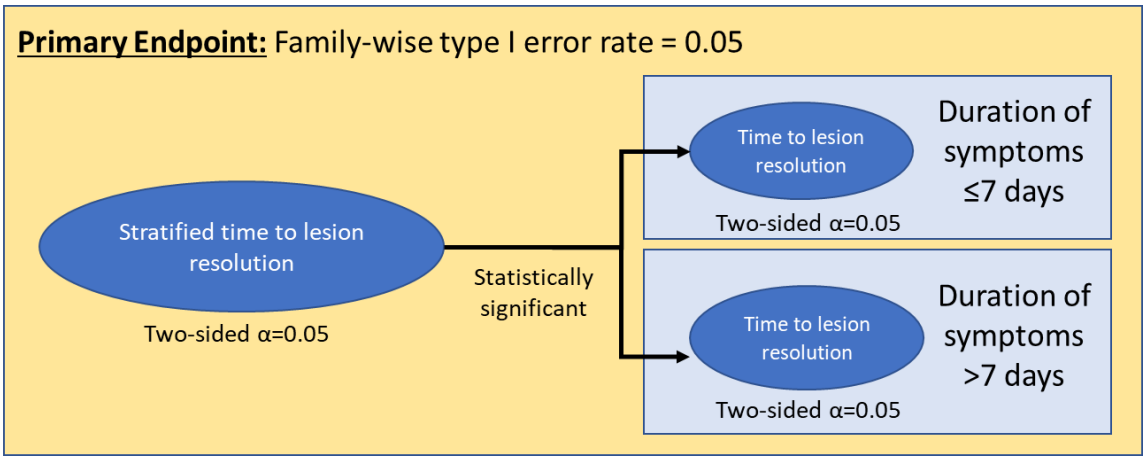
5.6.1 Hierarchical Testing of the Primary Endpoint

The final analysis will employ a hierarchical testing scheme described in Figure 1.

The primary endpoint will be assessed using Gray's test stratified by time from onset of prodromal symptoms (≤7 days or >7 days) with a two-sided type I error rate of α=0.05. The exact procedure for performing this test is described in Section 6.4 and Appendix 1. The first secondary endpoint plans to repeat this procedure within strata, but the assessment of statistical significance within strata will be conditional on the statisticial significance of the primary endpoint as follows.

If the primary stratified test is statistically significant, Gray's test will be performed again within each stratum, each test using a two-sided type I error rate of α=0.05. On the other hand, if the primary endpoint test is not statistically significant, no formal testing of time to lesion resolution within strata will be performed for the final analysis. In either case, confidence intervals for the treatment effect estimate overall and within strata will be reported.

Figure 1. Hierarchical testing scheme



5.7 Error Control Rate

The hierarchical testing procedure for the primary endpoint controls the family-wise type I error rate among the primary endpoint comparisons at level α=0.05 by the following argument:

1. The global null hypothesis for the primary endpoint is that there is no effect on time to lesion resolution in either symptom duration stratum. If this global null is true, then we would falsely declare an effect if and only if the stratified test results in a p-value less than or equal to 0.05. The probability of observing such a p-value under the global null is 0.05.
2. On the other hand, if the global null is false and there is an effect on time to lesion resolution then either a) both of the stratum-specific null hypotheses are false (the effect is present in both subgroups) or b) one of the stratum-specific null hypotheses is false (the effect is present only in one subgroup).
    a. If both stratum-specific null hypotheses are false, then it is not possible to make a type I error and thus the family-wise error rate is 0.
    b. If only one stratum-specific null hypothesis is false, then the probability of a type I error is the probability that the true stratum-specific null hypothesis is falsely rejected. This would require the p-value for that stratum to be less than or equal to 0.05, which occurs with probability 0.05.

5.8 Analysis populations

Efficacy analyses will be performed in the intent-to-treat (ITT) population, defined below, to form the basis for hypothesis testing described in Section 6. Safety analyses will be performed in the as-treated population. Sensitivity analyses may be performed in the as-treated or mITT populations or using as-randomized stratification to assess the impact of stratification errors.

In some cases, all exclusion criteria from certain analysis populations may not be anticipated in advance. For example, the actual study product received may not be immediately obvious if a participant receives multiple doses of the wrong product in error. If there are any such ambiguous cases, blinded review will occur to determine the proper handling of each case and sensitivity analyses may be performed to address the impact of the handling decision.

5.8.1 As-treated population

The as-treated population will include all participants who receive at least one partial or full daily dose of tecovirimat or placebo (i.e., any participant who receives study product). As-treated participants will be classified by the actual study product received (tecovirimat or placebo, which may not align with the arm dictated by the randomization) and actual number of days from onset of symptoms to randomization (≤7 days or >7 days).

- Any participants randomized to placebo who inadvertently receive tecovirimat for all doses will be grouped with the tecovirimat arm, and vice versa.

5.8.2 Intent-to-treat (ITT) population

The ITT population will include all randomized participants. ITT subjects will be classified by their randomized treatment arm and actual number of days from onset of symptoms to randomization (≤7 days or >7 days).

5.8.3 Modified Intent-to-treat (mITT) population

The mITT population will include all randomized participants with laboratory-confirmed mpox infection as defined in the protocol. Any participants who are randomized but later found to have been enrolled based on a false positive result (e.g., due to lab error) will be excluded from the mITT population. The mITT population will be classified based on their randomized treatment assignment and actual stratum (≤7 days or >7 days from symptom onset to randomization), which may not align with the stratum to which the participant was randomized.

5.9 Statistical software

Analyses will be performed in R version 4.0.0 or higher.

6 Descriptive and Inferential Analyses

This section describes the statistical analysis of each primary, secondary, and exploratory endpoint. It also describes planned statistical analyses and summaries of safety data. Tables and figures which will be produced by each primary and secondary analysis are described in the appropriate section. Example tables and figures for key analyses are presented in Section 10.

6.1 General Statistical Principles

Univariate summaries of continuous measures will include the mean, standard deviation, range, and 25th, 50th, and 75th percentiles. Categorical measures will be summarized using the number and percentage in each category.

For hazard ratio and subdistribution hazard ratio estimates, Wald confidence intervals will be used. Wilson's method for computing confidence intervals will be used when summarizing estimates of a single proportion and Mee's method will be used to compute confidence intervals for differences in proportions [8, 9]

6.2 Baseline Descriptive Statistics

Baseline values will be defined as the last observation prior to randomization.
Baseline characteristics including demographics, baseline comorbidities, prior medications, and baseline symptoms will be summarized by treatment arm in the ITT population. The proportion of missing observations for each variable will be provided.

6.3 Safety data

All reportable AEs will be recorded and coded by Medical Dictionary for Regulatory Activities (MedDRA). AEs will be presented by MedDRA system organ class (SOC) and preferred term (PT).

AEs on or after 1st dose date will be included in the tables. Listings will include all AEs, including those with onset prior to 1st dose.

For maximum by-grade AE tables, the number and proportion of participants with each AE will be tabulated by maximum severity, treatment arm, and overall. Participants with multiple occurrences of the same event will be counted once using the event with maximum severity.

For general AE tables (including but not limited to AEs, SAEs, AEs related to study treatment, AEs leading to treatment discontinuation, and AEs indicated as bacterial infections), the number and proportion of participants with each AE will be tabulated by treatment arm and overall. Participants with multiple occurrences of the same event will be counted once.

6.4 Primary Endpoint Analysis Details

The primary endpoint analysis will compare time to lesion resolution in the ITT population. A competing risks analysis will be performed to account for the competing risk of death on time to lesion resolution. There are two models of interest for a competing risks analysis: the cause-specific hazard model (which yeilds cause specific hazard ratios; csHR) and the subdistribution hazard model (which yields subdistribution hazard ratios; sHR) [10]. The cause-specific hazard function for a given event models the instantaneous rate of occurrence of that event in subjects who have not yet experienced any of the competing risks. The subdistribution hazard model, also referred to as a cumulative incidence function (CIF) regression model, describes the instantaneous rate of occurrence of the given event in subjects who have not yet experienced an event *of that type* [11]. The key difference is in the size of the risk set considered by the two approaches. In cause-specific hazard models a subject who experiences a competing event is censored at the time the competing

event occurred and therefore drops out of the risk set. In subdistribution hazard models an individual who experiences a competing event remains in the risk set and therefore experiences "immortal time" for the remainder of the period of observation – a time when they remain in the risk set yet cannot experience the event of interest (e.g., lesion resolution) because they have already experienced a competing event (e.g., death). Accepting that the risk set for the subdistribution hazards model includes individuals who are known to have zero probability of the event of interest has a critical implication—the magnitude of the effect of a covariate on the subdistribution hazard function is different from the magnitude of the effect of the covariate on the CIF [12]. Therefore, subdistribution hazard ratios describe only the direction of the association between a covariate and the CIF. Nevertheless, a test of statistical significance of the subdistribution hazard ratio (Gray's test, described below) does provide a test for the effect of tecovirimat on the cumulative incidence of the event of interest (lesion resolution) while accounting for the competing risk of death.

Subdistribution hazard ratios denote the magnitude of the relative change in the subdistribution hazard function associated with a one-unit change in the given covariate (e.g., treatment). Therefore, one is reporting the relative change in the instantaneous rate of the occurrence of the event in subjects who are event free or who have experienced a competing event. Cause-specific hazard ratios, on the other hand, denote the relative change in the instantaneous rate of the occurrence of the primary event in subjects who are currently event free. It has been suggested that competing risks analysis should report results from both cause-specific and subdistribution hazards model [13], which is the intent of the primary analysis.

The primary analysis of time to lesion resolution will be stratified by days from onset of prodromal symptoms, defined as the participant-reported date of the appearance of the first symptom) to randomization (≤7 vs >7 days). This stratification recognizes that antiviral effects are likely to be more apparent in the cohort who receives treatment more quickly after the onset of disease. Therefore, the primary analysis p-value will come from a stratified version of Gray's test [14]. Details for the implementation of this test in R are provided in Appendix 1.

Cause-specific and subdistribution hazard ratios for lesion resolution (with 95% confidence intervals) will be reported. Cumulative incidence of lesion resolution and death with accompanying 95% confidence intervals will be provided at 14, 21, and 28 days post-randomization.

Note that in scenarios where the competing risk (in this case, death) turns out to be rare (e.g., 0-3%), the subdistribution hazards model provides estimates of subdistribution hazard ratios for the primary event that are extremely similar to hazard ratios from a Cox proportional hazards model where those few who experience the competing event are treated as censored at the last possible observation time (i.e., for a beneficial primary event such as lesion resolution, assigning them the worst possible outcome under the model formulation). In the event that mortality is low, both analyses may be performed and it may be appropriate to express results as hazard ratios derived from a Cox model for the sake of familiarity to wider audiences. All Cox models implemented in analyses will use the Efron method of handling ties.

It is possible that a participant could achieve lesion resolution and subsequently die within the study observation period. In subdistribution hazard models where the primary event is lesion resolution, these individuals will be coded as having experienced the primary event only. The above coding specifications would mean individuals who died after achieving lesion resolution would be coded differently than individuals who died before. However, such a coding is a necessity for model fitting (both events cannot be observed, per the model definition) and this scenario is expected to be rare.

Unless there are substantial differences in mortality by arm (in which case time to lesion resolution is of less importance), this will have a minimal impact on interpretation of the subdistribution hazard ratio. A sensitivity analysis may be performed to assess this modelling choice.

It is also possible that participants could achieve lesion resolution and subsequently develop new lesions within the study period. For participants enrolled under protocol version 2.0 and prior (before the creation of the "sick visit" CRF), or for participants whose new lesions were observed at optional visits where no lesion assessments were performed (e.g., Day 59 or returning to the hospital post-discharge for an unscheduled visit), there may be no data to confirm the development of new lesions due to the absence of a data collection mechanism. It is important that the primary analysis be reproducible using the clinical database. Therefore, the primary analysis will use the following rules to determine outcomes.

1. Participants whose baseline lesions do not resolve by Day 29 will be censored at either the study day of their last performed lesion assessment or actual Day 29 (i.e., 28 days after randomization, not necessarily the visit date of Day 29), whichever comes first.

2. For participants whose baseline lesions resolve on or before actual Day 29, time to resolution is defined as the number of days from randomization to lesion resolution. Resolutions that occur after Day 29 will be censored at Day 29.

3. If participants develop new *confirmed mpox* lesions after resolution of baseline lesions but prior to the end of the Day 29 visit window (Day 36), there must be documentation of resolution of the new lesions by Day 29 in order for the primary endpoint value to be a lesion resolution event. Participants without documented resolution of the new lesions by Day 29 will be censored at the earliest of the study day of their last performed lesion assessment and actual Day 29. By default, this means participants with new confirmed mpox lesions between Day 29 and Day 36 will be censored at Day 29 (since those lesions cannot resolve by Day 29). To be *confirmed mpox* lesions, lesions must:

    a. Be documented on study CRFs (e.g., via a nonzero lesion count during a sick visit or regular visit lesion assessment)
    b. Be accompanied by a generic or clade I positive PCR result observed for any sample type during the period in which new lesions are observed.

A blinded adjudication committee will guide a sensitivity analysis for the primary endpoint that will incorporate site PI reports of recrudescent cases for which no data was recorded. The committee will be responsible for reviewing participant cases where study data or PI reports indicate possible recrudescent disease. Details of the adjudication committee membership and deliberations will be recorded in an adjudication committee report. The committee will remain blinded for all determinations, which should be completed before trial unblinding.

Other sensitivity analyses may be used to investigate the impact of the rules above.

All tests performed will be two-sided tests with α=0.05. A tabular summary of the number of events and number of participants censored (for each type of censoring) will also be provided.

Table: By arm overall and within strata: n per arm, cause-specific and subdistribution hazard ratios for lesion resolution with 95% CIs, and cumulative incidence of lesion resolution and death with 95% CIs at 14-, 21-, and 28-days post-randomization (Example Table 1)

Table: By arm overall and within strata: number and proportion of participants with an event (lesion resolution) and number and proportion censored (from each possible censoring type).

Figure: Cumulative incidence functions for lesion resolution and death by arm overall and within strata (Example Figure 1)

6.5 Secondary Endpoint Analysis Details

6.5.1 Time to lesion resolution, as defined in the primary endpoint, according to stratification by time from onset of illness as ≤ 7 days or > 7 days

Within-stratum analyses of time to lesion resolution will also be performed (by days from onset of symptoms ≤7 days vs >7 days). Cause-specific and subdistribution hazard ratios for lesion resolution will be presented with accompanying 95% confidence intervals. If the primary endpoint (stratified Gray's test of time to lesion resolution) is statistically significant, p-values from within-stratum hypothesis tests using Gray's test on the subdistribution hazard ratio for lesion resolution will additionally be reported. The within-stratum tests will use the same definitions of event times, censoring times, and coding for competing risks due to death.

6.5.2 Proportions with negative blood, oropharyngeal swab, and lesion swab PCR results 14 days after randomization

Available PCR data for a given participant, sample type, and timepoint are made up of 4 cycle threshold (CT) values: a clade I CT value, a clade II CT value, a generic orthopox CT value, and a positive control CT value. For each target, a CT value less than or equal to 40 indicates a positive result. No cases of clade II disease are expected to enroll in the trial. Therefore, the clade I and generic orthopox CT values are of most interest.

Little is known about the typical viral load trajectories of mpox patients in this setting. Exploratory work may be required to determine the most appropriate methods of data handling, for example to gain an understanding of how to handle cases where the clade I result is negative but the generic orthopox result is positive. Exploratory analyses will be blinded to study arm and will be described in detail in an appendix to the SAP.

The proportion of participants with negative PCR results 14 days after randomization will be summarized by arm for three compartments: blood, oropharyngeal swab, and lesion swab. Results will be described overall and within stratum (days from onset of symptoms to randomization ≤7 vs >7) and will include 95% confidence intervals.

A positive result on at least one of the three compartments is required for study participation, but it is possible that some participants may be PCR negative at baseline in up to two compartments. Denominators for proportions will be the number who were positive at baseline for that compartment.

For participants who are missing Day 15 PCR data, a last-observation-carried-forward approach will be taken to impute Day 15 positive/negative status.

For participants who achieve lesion resolution prior to Day 15, clinical judgement may determine that no further lesion swab samples should be collected (since this would require opening the already-resolved lesions). To account for this possibility, participants who achieve lesion resolution prior to Day 15 will be considered negative at Day 15 for the lesion swab PCR compartment, unless Day 15 data exist.

6.5.3 Mortality

Per-arm mortality (n died by Day 29/N in group and percentage) will be summarized overall, by site, by stratum (days from onset of symptoms to randomization ≤7 vs >7), and in site/stratum combinations. Mortality is considered both an efficacy and a safety analysis and will be summarized using both the ITT and the as-treated populations. .

Table: Overall and by stratum: n per arm, n died by Day 29, percentage of deaths (Example Table 2).

6.5.4 Time to death

Time to death will be analyzed using a standard survival analysis approach. A Cox proportional hazards model will be constructed using death as the outcome of interest and censoring at Day 29 any individual who survives and completes a Day 29 visit [15]. Individuals who leave the study early and do not complete a Day 29 visit will be censored at their last observation time. Separate models will be created in the overall ITT population (where the model will be stratified) and by stratum (days of symptoms prior to randomization ≤7 or >7). For all models, ties will be handled using the Efron technique [16, 17]. Estimates of the hazard ratio and 95% CIs will be provided. No formal hypothesis testing is planned using these models. Kaplan-Meier figures of the probability of survival to 28 days post-randomization will be provided.

Table: Overall and by stratum: HR (95% CI) for 28-day mortality on the tecovirimat arm relative to the placebo arm.

Figure: Kaplan-Meier survival curve in the overall population and within strata.

6.5.5 Frequency and duration of clinical symptoms

A solicited symptom evaluation will be performed for all participants at baseline and during each daily follow-up visit as well as at Days 29 and 59. Overall and by stratum, frequency of symptoms by arm (n/N in risk set, %) will be provided for each symptom at baseline and Days 8, 15, 29, and 59 using the ITT population.

For each participant, duration of clinical symptoms will be defined as the number of days from participant-reported onset of symptoms until the day when: 1) the participant is recorded as experiencing no solicited symptoms and 2) the participant is not recorded as experiencing any solicited symptoms on subsequent days. For participants who are discharged without having met this definition (i.e., those who are discharged with at least one solicited symptom still present), the assumed first day without any symptoms will be the day after discharge. Missing symptom data will be handled using a last-observation-carried-forward (LOCF) approach: symptoms will be imputed as present if the most recent non-missing value was present and not present if the most recent non-

missing value was not present or if there is no non-missing data for that symptom. Duration of clinical symptoms will be summarized as a continuous measure by stratum.

Table: Overall and by stratum: number and proportion of participants with each solicited symptom out of the number with data. Separate tables will be presented for baseline and Days 8, 15, 29, and 59.

Table: For each arm: duration of clinical symptoms by stratum (n, mean, SD, minimum, 25th percentile, median, 75th percentile, maximum)

6.5.6 Incidence of SAEs, AEs requiring drug discontinuation, and other AEs

All adverse events (AEs) will be graded by severity (1-5) and assessed for the likelihood that the event was caused by the study agent (definitely related, probably related, possibly related, unlikely related, or not related). AEs will also be coded using the current version of the Medical Dictionary for Regulatory Activities (MedDRA).

AE summaries will be presented in the as-treated population according to MedDRA System Organ Class (SOC), MedDRA Preferred Term (PT), severity, and causality. Summaries will include the number and proportion of participants who experienced an AE with that SOC, PT, severity, or causality. Summaries of SAEs and AEs requiring drug discontinuation will be provided separately using the same approach.

At a minimum, number and percentage of participants experiencing 1 or more and 2 or more events for each of the following categories will be reported: AEs, related AEs, AEs requiring drug discontinuation, grade 3+ AEs, related grade 3+ AEs, SAEs, related SAEs. Per arm differences in proportions with 95% confidence intervals will be provided.

Listings will be prepared of all AEs, AEs requiring drug discontinuation, and all SAEs.

Table: For all SOCs/PTs/severity grades/causality assessments: n with an event of that type, N in population, percentage.

Table: Number and percentage of participants experiencing 1 or more and 2 or more events for each of the following categories: AEs, related AEs, AEs requiring drug discontinuation, grade 3+ AEs, related grade 3+ AEs, SAEs, related SAEs; per arm differences in proportions with 95% confidence intervals

Listings: All AEs, All AEs requiring drug discontinuation, All SAEs

6.5.7 Incidence of bacterial infections

Bacterial infection is an adverse event of specific interest. Incidence of bacterial infections by study arm will be summarized by the number and percentage of participants by study arm who experience at least one bacterial infection AE, overall and by stratum in the as-treated population. Differences in proportions with 95% confidence intervals will be reported.

Table: Number and percentage of participants with at least one bacterial infection adverse event, difference in proportions with 95% confidence interval

6.5.8 Automated image analysis of lesion counts and characteristics over time

An automated image analysis algorithm will be trained using longitudinal photographs of the lesion assessment region. Lesion counts and characteristics from photographs will be assessed and compared to evaluate the performance of the algorithm relative to a human rater. Separate documentation of the details of the algorithm will be prepared. Depending on the performance of the algorithm, several additional analyses might be of interest. These analyses would be described in separate documentation and no SAP update is planned to detail these exploratory procedures.

6.6 Exploratory Endpoint Analysis Details

Exploratory analyses are described in some detail here, but the SAP does not intend to prespecify all analyses, tables, or figures that may be deemed appropriate to address each exploratory endpoint. Analyses or summaries described below represent the minimum that will be reported.

6.6.1 Frequency and location of persistent residual lesions

The study intends to collect baseline data on the total number of bodily lesions and longitudinal data on lesion counts for an assessment region comprised of the right arm and right leg. Preliminary data analysis using data from an observational cohort indicates that lesion resolution in the assessment area will closely coincide with full body resolution. However, this may not necessarily be the case in the present study. Therefore, after lesions in the assessment region are entirely resolved, data collection will continue to entail daily assessments of whether any lesions persist in other areas: non-assessment-region arm, non-assessment region leg, head, front of trunk, back of trunk, genitals, and other. This exploratory analysis will describe any areas recorded as having unresolved lesions after lesions in the assessment region have all resolved.

The overall number and percentage of participants who experience any persistent residual lesions and who experience persistent lesions in each body area will be reported.

6.6.2 Lesion progression over the study period

Lesion progression over the study period will be compared using descriptive analyses of lesion count trajectories in the assessment region. Lesion counts will be summarized as a continuous variable at baseline/Day 1 and at Days 3, 5, 8, 15, 29. Lesion counts will be visualized using a figure displaying median and IQR of lesion counts by study arm at each timepoint.

The correlation between total number of lesions at baseline and number of assessment region lesion at baseline and Days 3, 5, 8, and 15 will be reported.

Additional analyses of lesion progression may include a derived variable analysis. For example, simple linear regression may be used to estimate a slope of the lesion count trajectory for each participant over the first X days. These slopes may be used for simple comparison (e.g., of

confidence intervals for the slope comparing participants on each treatment arm) or in more sophisticated modeling (e.g., used as covariates in a prediction model).

6.6.3 Exposure history

Participants' exposure history for MPX will be collected at baseline and will include whether the participant handled uncooked meat, handled wild animals that may have been infected with MPX, came into direct contact with a person suspected to have MPX, or had other potential exposure risks. The number and proportion of participants self-reporting their status for each of the exposure risks collected will be reported.

6.6.4 Associations between baseline disease severity and outcomes

The goal of this exploratory endpoint is to develop a metric for baseline mpx disease severity that is predictive of future clinical outcomes. Models will be implemented to assess the effect of relevant baseline characteristics (e.g., age, sex, baseline lesion count, selected comorbidities) on clinical outcomes and a predictive model will be developed using a training partition of the final study data. The final model will be evaluated using a testing partition of the final study data and results will be reported. The model building exercise will be documented elsewhere and results from any model would ideally be externally validated as part of future work.

6.6.5 Differential outcomes according to presence of anti-OPXV antibodies

This analysis, though given distinction in the study protocol as an exploratory endpoint, will be handled as a subgroup analysis described in Section 7.1. Analyses of the primary and secondary endpoints may be repeated in subgroups of participants with and without presence of anti-OPXV antibodies.

6.6.6 Viral persistence

Available PCR data for a given participant, sample type, and timepoint are made up of 4 cycle threshold (CT) values: a clade I CT value, a clade II CT value, a generic orthopox CT value, and a positive control. For each target, a CT value less than or equal to 40 indicates a positive result. No cases of clade II disease are expected to enroll in the trial. Therefore, the clade I and generic orthopox CT values are of most interest. Questions remain over the proper handling of cases where the clade I result is negative but the generic orthopox result is positive. Analyses of PCR data will specify whether such results were considered positive or negative and may be performed both ways.

Starting on Day 1, participants will receive PCR tests of the blood, skin lesions, and oropharyngeal swabs every other day until two consecutive negative results are observed (separately for each compartment).

Figures will be produced showing the proportion of participants who are PCR negative according to blood sample results, oropharyngeal (OP) swabs, and lesion swabs by study arm through Day 29.

Median and IQR viral load for each compartment will be shown longitudinally through Day 29.

A set of three competing risks analyses will be performed in the mITT population where time to first compartment-specific PCR negative result is the event of interest and death is the competing event. Definitions of censoring and handling for individuals who die within the study period but after achieving PCR negativity (expected to be rare) will be identical to the primary endpoint analysis. In theory, the protocol specifies that consecutive negative results from blood PCR is a requirement for discharge. In practice, participants may be discharged who did not meet this requirement. This analysis is fairly strict on this point: participants who are discharged without a negative result will be considered censored rather than being counted as events, even if the Day 29 blood PCR sample result is negative.

Cause-specific and subdistribution hazard ratios for PCR negativity in each compartment will be reported with 95% confidence intervals. Cumulative incidence of PCR negativity and accompanying 95% confidence intervals will be provided at 14-, 21-, and 28-days post-randomization.

If deaths are rare, the competing risks analysis approach may not be strictly needed. A standard survival analysis using a Cox proportional hazards model may also be implemented and the results compared to the competing risks analysis for context.

Note that in the original (V1.0) study protocol, time to first negative blood PCR result was listed as a key secondary endpoint and formed part of a planned hierarchical testing scheme in which time to blood negativity would be tested formally using a stratified version of Gray's test at level 0.05. The procedure would then continue to testing the mortality secondary endpoint using Boschloo's test if the stratified Gray's test of time to blood PCR negativity were statistically significant. This testing plan brought the overall type I error rate for all planned tests for the study to at most 0.10. However, early baseline data indicated that a smaller-than-expected proportion of participants were PCR positive in the blood at baseline (54/87 enrolled and with data as of 4 June 2023; 62.1%). Recognizing that time to blood negativity may no longer be a meaningful outcome for a large group of participants, a decision was made by the blinded statistical team to move time to PCR negativity (for blood and all other compartments) to an exploratory endpoint and reduce the planned formal hypothesis testing to the primary endpoint only, which also has the beneficial effect of reducing the overall type 1 error rate across all planned tests to 0.05.

Table: For each compartment by arm overall and within strata (days from onset, as in the primary analysis): n per arm, cause-specific and sub-distribution hazard ratios for time to PCR negativity with 95% CIs, cumulative incidence of PCR negativity with 95% CIs at 14-, 21-, and 28-days post-randomization (Example Table 1)

Figure: Cumulative incidence functions for PCR negativity and death by arm overall and within strata (Example Figure 1)

6.6.7 Antibody titers

Blood for storage is planned to be obtained from all participants pre-dose on Days 1, 7, 13, 29, and 59. Geometric mean and geometric standard deviation by arm will be reported for all available timepoints. Missing data will be reported by timepoint for each arm. Descriptive statistics will be reported with all missing data omitted and with a LOCF approach implemented to impute missing post-baseline data.

6.6.8 Viral Sequencing

If possible, viral sequencing facilitate a descriptive analysis of sequencing differences between virus isolated from participants from different geographic areas and with differing clinical trajectories and responses to tecovirimat as well as between virus isolated prior to and after treatment.

6.6.9 Pharmacokinetic assessment

A basic pharmacokinetic assessment is included in the study to address whether the dose concentrations of tecovirimat administered in field conditions in rural DRC align with what is expected based on previous work. Pre-dose concentration of tecovirimat in blood measured on Day 7 (e.g., representing the status of the participant after 6 full days of doses have been received) will be summarized as a continuous variable by stratum and by quantiles of age and weight.

6.6.10 Incidence of Recrudescent Disease

A descriptive summary (including the rate and clinical characteristics) of observed cases of recrudescent disease, as defined in the protocol, will be provided.

7 Additional Analyses

7.1 Subgroup Analyses

A planned set of subgroup analyses will evaluate the treatment effect as measured by the primary endpoint across a series of subgroups. A forest plot will display the point estimate for the sub-distribution hazard ratio for lesion resolution from the primary (stratified) analysis and in the following subgroups:

- Duration of symptoms prior to randomization (≤7 days or >7 days),
- Study site (if more than one site enrolls participants),
- Age,
- Sex (male/female),
- Baseline lesion burden according to a scale developed for smallpox by WHO (Table 2),
- Baseline viral load (≤ or > median in the overall sample),
- HIV status,
- CD4 count (if data is available; quartiles among the HIV-positive subgroup),
- Virus clade (if clade determination is possible), and
- Presence/absence of anti-OPXV antibodies at baseline.

A forest plot will display point estimates and 95% confidence intervals for the lesion resolution rate ratio by subgroup. Interaction tests may be conducted to determine whether the effect of treatment varies by subgroup. Additional subgroups may be identified and evaluated in addition to those listed here.

Table 1. WHO smallpox baseline lesion severity categories

WHO smallpox baseline lesion severity category	Number of total bodily lesions at baseline
Subclinical	0 – 4

Mild	5 – 25
Moderate	26 – 100
Severe	101 – 250
Grave	>250

8 Interim Analysis Plan

A Data and Safety Monitoring Board (DSMB) will monitor ongoing results to ensure the well-being and safety of participants as well as study integrity. The DSMB will recommend stopping the study early for efficacy only when there is substantial evidence of a treatment benefit. Similarly, the DSMB will recommend stopping early for futility only when there is substantial evidence to make such a recommendation.

8.1 Interim Safety Review

A blinded pharmacovigilance committee will conduct safety oversight for this study. A blinded medical monitor in the DRC has been appointed from the pharmacovigilance committee to be responsible for performing oversight and review of safety assessments. In addition, the DSMB will perform safety data reviews approximately every 6 months while the study is ongoing. The DSMB may request to be unblinded for these reviews.

8.1.1 Safety Monitoring by the Pharmacovigilance Committee

Safety data including but not limited to adverse event listings and tabulations of AEs and SAEs will be prepared by the unblinded statistical team and made available to the medical monitor and pharmacovigilance committee via a secure folder on the study website.

The pharmacovigilance committee and medical monitor are intended to remain blinded. However, the medical monitor can request unblinding in certain scenarios as defined in the protocol.

The pharmacovigilance committee will also receive reports of all deaths, SAEs, unanticipated problems (UPs), and pregnancies.

8.1.2 Safety Monitoring by the DSMB

The DSMB will receive reports of all deaths, SAEs, unanticipated problems (UPs), and pregnancies.

DSMB safety reviews are planned to occur approximately every 6 months while the study is ongoing. At these reviews, the DSMB will receive safety data including but not limited to adverse event listings and tabulations of AEs, AEs requiring drug discontinuation, and SAEs. Additional data summaries such as longitudinal vital signs, laboratory values, and symptom frequency may also be provided.

8.2 Interim Efficacy and Futility Review

NOTE: The text below describes the original interim monitoring plan. See Appendix 2 for the updated interim monitoring plan.

Interim efficacy analyses are planned when the study reaches 1/3 and 2/3 of the 318 lesion resolution events specified by the sample size calculation (i.e., at 1/3 and 2/3 information time). Therefore, the interim analyses are planned to be conducted after 106 and 212 events have been observed. The unblinded statistical team will monitor the number of lesion resolution events and will suggest a data cutoff date prior to each interim analysis designed to come as close as possible to the prespecified information time. The statistical team will then coordinate with the DSMB executive secretary to arrange a DSMB meeting.

The actual information time of each interim analysis may differ slightly from what is planned. The true information fraction at an interim analysis will be computed as t=n/318, where n is the number of lesion resolution events observed at the time of the data cutoff.

The Lan-DeMets spending function analog of the O'Brien-Fleming boundaries will be used to monitor the primary endpoint as a guide for the DSMB for an overall two-sided type-I error rate of 0.05 [18]. Two one-sided boundaries will be constructed at level α=0.025. The R package *ldbounds* can be used to calculate boundaries. Specifically, if the interim analyses occur exactly at the planned information time, the Z-score boundaries for statistical significance are Z=+/-3.7103 at the first interim (1/3 information), Z=+/-2.5112 at the second interim (2/3 information), and Z=+/-1.994 at the third interim.

Conditional power will be presented as an additional guide to the DSMB at interim analyses for assessment of futility. Conditional power is the probability of obtaining a statistically significant result by the end of the trial given the data accumulated thus far, assuming a hypothesized treatment effect.

Let t be the current information time, Z_t be the current Z value at time t, and θ_F be the log of the sub-distribution hazard ratio assumed for the distribution of future data. Conditional power is obtained as $CP(t, \theta_F)= \Phi\{Z_{CP}(t, \theta_F)\}$ where Φ is the standard normal distribution function and

$$Z_{CP}(t, \theta_F) = [\, Z_t(t^{1/2}) + \theta_F(1-t) - \Phi^{-1}\{1-\alpha/2\}\,] \,/\, [\, (1-t)^{1/2}\,],$$

as described by Lachin [19].

The primary conditional power calculation will be calculated under the originally hypothesized sub-distribution hazard ratio of 1.40 ("under the design"; θ_F =log(1.40)). A secondary conditional power calculation will use the observed sub-distribution hazard ratio at the time of the interim analysis ("under the trend"; θ_F is the current log sub-distribution hazard ratio for the stratified test of the primary endpoint).

If conditional power under the originally hypothesized effect is less than 20%, consideration should be given to stopping the trial, though this is not a requirement. Conditional power computed using the current trend will be provided for context only since that estimate is more variable.

The unblinded statistical team will prepare closed reports for DSMB reviews containing a clear description of the current test statistic for the primary endpoint as it relates to crossing prespecified boundaries, the conditional power of the study as described above, and any other relevant information needed to monitor the trial. The DSMB may suggest additional summaries to be provided. Analyses in closed reports will be presented with masked treatment codes to protect against the possibility that the DSMB report may fall into the wrong hands. The DSMB may formally request the unblinding of the masked treatment codes at any time. Open reports of study progress

will also be prepared for review by a wider, blinded audience and will contain overall summaries across both treatment arms.

9 References

[1] Rimoin AW, Mulembakani PM, Johnston SC, et al. Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. Proceedings of the National Academy of Sciences. 2010;107(37):16262-16267

[2] Parker S, Handley L, Buller RM. Therapeutic and prophylactic drugs to treat orthopoxvirus infections. Future Virol. Nov 2008;3(6):595-612. doi:10.2217/17460794.3.6.595

[3] Mucker EM, Goff AJ, Shamblin JD, et al. Efficacy of tecovirimat (ST-246) in nonhuman primates infected with variola virus (Smallpox). Antimicrob Agents Chemother. Dec 2013;57(12):6246-53. doi:10.1128/AAC.00977-13

[4] SIGA Technologies Inc. TPOXX (tecovirimat) [full US prescribing information]. 05/2022. Accessed 11 August, 2022. https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/214518s000lbl.pdf

[5] Pittman, PR, et al. Clinical characterization of human monkeypox infections in the Democratic Republic of the Congo. medRxiv 2022.05.26.22273379. https://doi.org/10.1101/2022.05.26.22273379

[6] Tai, BC., Wee, J. & Machin, D. Analysis and design of randomised clinical trials involving competing risks endpoints. Trials 12, 127 (2011). https://doi.org/10.1186/1745-6215-12-127

[7] Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics. 1983 Jun;39(2):499-503. PMID: 6354290.

[8] Wilson, EB (1927) Probable Inference, the Law of Succession, and Statistical Inference, Journal of the American Statistical Association, 22:158, 209-212, DOI: 10.1080/01621459.1927.10502953

[9] Mee, R. W., & Anbar, D. (1984). Confidence bounds for the difference between two probabilities. *Biometrics*, 1175-1176.

[10] Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016 Feb 9;133(6):601-9. doi: 10.1161/CIRCULATIONAHA.115.017719. PMID: 26858290; PMCID: PMC4741409.

[11] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association. 1999;94:496–509.

[12] Austin, PC, Fine, JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in Medicine. 2017; 36: 4391– 4400. https://doi.org/10.1002/sim.7501

[13] Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J Clin Epidemiol. 2013 Jun;66(6):648-53. doi: 10.1016/j.jclinepi.2012.09.017. Epub 2013 Feb 14. PMID: 23415868.

[14] Zhou B, Latouche A, Rocha V, Fine J. Competing risks regression for stratified data. Biometrics. 2011 Jun;67(2):661-70. doi: 10.1111/j.1541-0420.2010.01493.x. Epub 2010 Dec 14. PMID: 21155744; PMCID: PMC3431205.

[15] Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B (Methodological), 34(2), 187–220. http://www.jstor.org/stable/2985181

[16] Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. Journal of the American Statistical Association, 72(359), 557–565. https://doi.org/10.2307/2286217

[17] Hertz-Picciotto I, Rockhill B. Validity and efficiency of approximation methods for tied survival times in Cox regression. Biometrics. 1997 Sep;53(3):1151-6. PMID: 9333345.

[18] DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med. 1994 Jul 15-30;13(13-14):1341-52; discussion 1353-6. doi: 10.1002/sim.4780131308. PMID: 7973215.

[19] Lachin JM. A review of methods for futility stopping based on conditional power. Stat Med 2005;24:2747-2764.

10 Example Tables and Figures

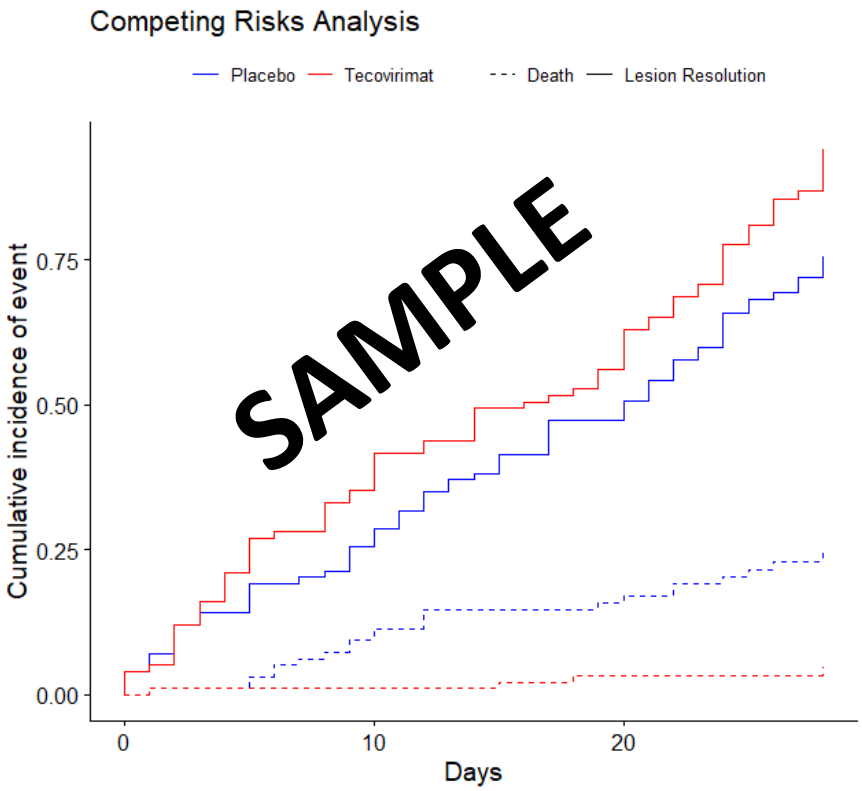
Example Table 1. Presentation of Competing Risks Analysis Results

	Overall (Stratified by Duration of Onset of Symptom)		Days from symptom onset to randomization			
			≤7 days		>7 days	
	Placebo (N=)	Tecovirimat (N=)	Placebo (N=)	Tecovirimat (N=)	Placebo (N=)	Tecovirimat (N=)
[Primary event]						
N (%) achieving [primary event] within 28 days	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Cause-specific hazard ratio (95% CI)	csHR (95% CI)		csHR (95% CI)		csHR (95% CI)	
Subdistribution hazard ratio (95% CI)	sHR (95% CI); p-value as applicable		sHR (95% CI); p-value as applicable		sHR (95% CI); p-value as applicable	
Cumulative Incidence (95% CI) of [primary event]						
14 days						
21 days						
28 days						
[Competing event]						
N (%) achieving [competing event] within 28 days	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Cause-specific hazard ratio (95% CI)	csHR (95% CI)		csHR (95% CI)		csHR (95% CI)	
Subdistribution hazard ratio (95% CI)	sHR (95% CI)		sHR (95% CI)		sHR (95% CI)	
Cumulative Incidence (95% CI) of [competing event]						
14 days						
21 days						
28 days						

Example Table 2. Presentation of Mortality Results

	Overall		Days from symptom onset to randomization			
			≤7 days		>7 days	
	Placebo (N=)	Tecovirimat (N=)	Placebo (N=)	Tecovirimat (N=)	Placebo (N=)	Tecovirimat (N=)
28-day Mortality						
N (%) died by Day 29	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)

Example Figure 1. Cumulative Incidence of Lesion Resolution and Death



Appendix 1 – Competing Risks Analyses in R

See attached R file.

Appendix 2 – Sample Size and Interim Monitoring

As of protocol version 4.0 (the basis for this analysis plan), there have been two adjustments made to the planned sample size and target number of events for the trial.

The first, documented in protocol version 3.0, was based on the blinded sample size re-estimation procedure described in Section 5.4. The blinded sample size re-estimation committee met on 29 August 2023. The committee decisions are documented in the PALM007 Blinded Sample Size Re-Estimation Committee Summary report. The committee noted that enrollment was faster than expected and the primary event rate was substantially higher than anticipated. The committee recommended that the expected total number of events be revised to 440 rather than 318. Since this is an event driven trial, an expectation of more events led to a revised power calculation. Using the formula in Section 5.4, power to detect a 40% improvement in the lesion resolution rate ratio would improve from 85% (318 events) to 94% (440 events), despite the fact that the target sample size remained the same (450). The sample size re-estimation committee additionally recommended that the number of planned analyses be reduced to two, for reasons outlined in protocol version 3.0. The DSMB approved of this update, which was documented in protocol version 3.0, but then subsequently recommended (after the first interim analysis review) returning to the 3-total-analyses plan.

The second adjustment, documented in protocol version 4.0, increased the target sample size from 450 to 600 and officially returned to 3 planned analyses per the DSMB recommendation. The study team responsible for the sample size increase decision was blinded. The blinded decision was based on multiple motivating factors outlined in the study protocol. In short, it was felt that the study must aim to provide definitive evidence toward evaluation of the primary endpoint and several highly important secondary and exploratory endpoints. Absent definitive evidence of efficacy, futility, or harm (all of which are monitored by the DSMB), there is a scientific benefit in continuing to enroll participants while the outbreak continues and there remains enough study product to randomize and treat participants. An internal pharmacy inventory conducted during the planning of protocol version 4.0 estimated that there was enough study product available to enroll up to 600 participants. This target sample size is expected to provide approximately 550 lesion resolution events, which will provide excellent power (>90%) for the primary endpoint overall and within the largest of the two randomization strata (days from symptom onset to randomization ≤7 days, expected to be approximately 75% of participants).

Increasing the target number of events necessitates an update to the timing of planned interim monitoring. The initial plan for interim efficacy monitoring (Section 8.2) included 3 planned analyses at 1/3, 2/3, and full information time based on the original target of 318 events (i.e., after 106, 212, and 318 events). As described above, the original plan was revised to instead target 440 events and two total analyses in protocol v3.0, which was the protocol in effect at the time of the first interim analysis. Protocol v4.0 was approved prior to the second interim analysis (which has not yet taken place at the time of SAP v2.0 finalization). To provide information on how the first interim analysis statistical boundaries were computed, and to direct how the second interim analysis and final analysis thresholds *will be* computed, Table A1 describes the approach for all relevant protocol versions.

Table A1. Monitoring Plans by Protocol Version

	Information Time	Z Score Boundary	Nominal Alpha (p-value threshold for significance)
Protocol 1.0 and 2.0 N=450 Target 318 events	0.33 (106 events)	±3.71	0.0002
	0.67 (212 events)	±2.51	0.0120
	1 (318 events)	±1.99	0.0463
Protocol 3.0 N=450 Target 440 events	0.50 (220 events)	±2.96	0.0031
	1 (440 events)	±1.97	0.0490
Protocol 4.0 N=600 Target 550 events	0.40 (220 events)	±3.36	0.0008
	0.60 (330 events)	±2.68	0.0074
	1 (550 events)	±1.98	0.0476

Note that the nominal alpha reported in Table A1 is a function of the number and timing of the interim analyses. The blinded study team responsible for protocol updates and development of the statistical analysis plan were not informed of the number of events captured in the first interim efficacy analysis. It is assumed that 220 events were included in the first analysis review, but readers should be aware that the true information time of that analysis may differ and that this difference would also impact the nominal alpha for the final analysis slightly. This effect is minimal for small to moderate departures from 220 events. The number of events included in the second interim analysis is also an assumption. The effect of "overshooting" the second interim analysis by including more than 330 events can be accounted for when computing the boundaries, as can the fact that the first interim analysis technically "spent" more type I error than needed (0.0031 rather than 0.0008) since it occurred under protocol v3.0. The actual boundaries used for the analyses are computed by the unblinded statistics team and clearly stated in closed interim efficacy reports prepared for the DSMB.

Summary of changes to the SAP (v1.0 to v2.0) for PALM007

1. We clarified how we'll handle recrudescent cases (Section 6.4).
   The "new" rules:
   1. Participants whose baseline lesions do not resolve by Day 29 will be censored at either the day of their last performed lesion assessment or actual Day 29 (i.e., 28 days after randomization rather than occurring within the Day 29 visit window), whichever comes first.

   Note: This is the same as before.

   2. For participants whose baseline lesions resolve on or before actual Day 29, time to resolution is defined as the number of days from randomization to lesion resolution. Resolutions that occur after Day 29 will be censored at Day 29.

   Note: This is the same as before.

   3. If participants develop new *confirmed mpox* lesions after resolution of baseline lesions but prior to the end of the Day 29 visit window (Day 36), there must be documentation of resolution of the new lesions by Day 29 in order for the primary endpoint value to be a lesion resolution event. Participants without documented resolution of the new lesions by Day 29 will be censored at the earliest of the study day of their last performed lesion assessment and actual Day 29. By default, this means participants with new confirmed mpox lesions between Day 29 and Day 36 will be censored at Day 29 (since those lesions cannot resolve by Day 29). To be *confirmed mpox* lesions, lesions must:

      Be documented on study CRFs (e.g., via a nonzero lesion count during a sick visit or regular visit lesion assessment), *and*
      Be accompanied by a generic or clade I positive PCR result observed for any sample type during the period in which new lesions are observed.

   Note: This is new and applies only to individuals with evidence of possible recrudescence. The adjudication committee will review them all. The rule describes how they'll be handled for the primary analysis (based on data). The committee report describes how they'll be handled for a sensitivity analysis (using the committee determinations, sometimes based on PI reports only).

2. We clarified how we'll handle recrudescent cases (Section 6.4), describing how the adjudication committee decisions will be used in data analyses.

   A blinded adjudication committee will guide a sensitivity analysis for the primary endpoint that will incorporate site PI reports of recrudescent cases for which no data was recorded. The committee will be responsible for reviewing participant cases where study data or PI reports indicate possible recrudescent disease. Details of the adjudication committee membership and deliberations will be recorded in an adjudication committee report. The committee will remain blinded for all determinations which will be completed before trial unblinding.

3. We added a statement on PCR result interpretation.
   Available PCR data for a given participant, sample type, and timepoint are made up of 4 cycle threshold (CT) values: a clade I CT value, a clade II CT value, a generic orthopox CT value, and a positive control. For each target, a CT value less than or equal to 40 indicates a positive result. No cases of clade II disease are expected to enroll in the trial. Therefore, the clade I and generic orthopox CT values are of most interest.

4. We added Appendix 2 to provide calculations related to sample size increases.