

STUDY PROTOCOL

Open Access



# Statistical analysis plan for the “empirical treatment against cytomegalovirus and tuberculosis in HIV-infected infants with severe pneumonia” clinical trial

Sara Domínguez-Rodríguez<sup>1,2\*</sup> , David Lora<sup>1,3,4</sup>, Alfredo Tagarro<sup>1,2,5</sup>, Cinta Moraleda<sup>1,6</sup>, Álvaro Ballesteros<sup>1</sup>, Lola Madrid<sup>1,7</sup>, Lilit Manukyan<sup>1</sup>, Olivier Marcy<sup>8</sup>, Valeriane Leroy<sup>9</sup>, Alessandra Nardone<sup>10</sup>, David Burger<sup>11</sup>, Quique Bassat<sup>12,13,14,15,16</sup>, Matthew Bates<sup>17</sup>, Raoul Moh<sup>18</sup>, Pui-Ying Iroh Tam<sup>19,20</sup>, Tisungane Mvalo<sup>21</sup>, Justina Magallhaes<sup>15</sup>, W. Chris Buck<sup>22,23</sup>, Jahit Sacarlal<sup>23</sup>, Victor Mussime<sup>24,25</sup>, Chishala Chabala<sup>26</sup>, Hilda Angela Mujuru<sup>27</sup>, Pablo Rojo<sup>3,6,28</sup> and on behalf of EMPIRICAL group

## Abstract

**Background** The EMPIRICAL trial aims to assess safety and efficacy of an empirical treatment against cytomegalovirus (CMV) and tuberculosis (TB) compared to standard of care (SoC), on adverse events and 15-day and 1-year mortality among infants living with HIV hospitalized with severe pneumonia in Africa.

**Methods and design** The EMPIRICAL trial (NCT03915366) is an international multicenter phase II-III, open-label randomized factorial clinical trial conducted in six African countries. The trial has four randomization arms in a 1:1:1:1 fashion with patients allocated to (i) TB-Treatment plus SoC, (ii) valganciclovir plus SoC, (iii) both TB-Treatment and valganciclovir plus SoC, and (iv) SoC only.

**Discussion** This paper describes the statistical analysis plan (SAP) for the trial which, per the study publication plan, needs to be published prior to the database lock and final analysis results. The SAP includes details of the analyses to be undertaken and unpopulated tables that will be reported to address primary and secondary endpoints. The database will be locked on 31st January 2025.

**Trial registration** ClinicalTrials.gov: NCT03915366 (registered on April 16, 2019), Universal Trial Number: U111-1231-4736, Pan African Clinical Trial Registry: PACTR201994797961340.

**Keywords** Statistical analysis plan, Child mortality, Cytomegalovirus, Empirical, Factorial, Factorial randomized clinical trial, HIV, Infants, Pneumonia, Tuberculosis, Valganciclovir

\*Correspondence:

Sara Domínguez-Rodríguez  
sara.dominguez.r@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Pneumonia remains the main cause of death in children in the postnatal period. Although underdiagnosed, tuberculosis is also a common cause of death, particularly in children with HIV infection. In 2022, according to UNAIDS, 1.5 million children under 15 years of age were living with HIV worldwide, mainly in Sub-Saharan Africa, and the global number of child deaths is 110,000 to 120,000 per year. Of those, 40,000 deaths could be attributable to tuberculosis (TB), despite early initiation of antiretroviral treatment (ART) [1, 2].

The current World Health Organization (WHO) guidelines include a standard of care (SoC) intervention to treat pneumonia in children living with HIV (CLHIV) caused by *Pneumocystis jirovecii*, *Streptococcus pneumoniae*, and *Haemophilus influenzae* b [3]. This intervention has insufficiently decreased mortality, allegedly due to other important causes of death, such as cytomegalovirus (CMV) and TB. Both are still underdiagnosed and undertreated in this population [4] due to the barriers and difficulties in diagnosis. The current SoC includes microbiological testing for TB in those children with suspected TB. However, even with this approach, a significant number of TB patients remain undiagnosed or late diagnosed. According to the WHO, improving access to oral systemic treatment with valganciclovir to treat CMV pneumonia is an explicit priority in children according to the 2017 Advanced HIV Disease Guidelines [5].

Systematic empirical treatment for CMV and TB may be lifesaving for infants living with HIV. Systematic empirical treatment for TB in severely immunosuppressed HIV-infected patients without evidence of active TB disease is an open question that has been assessed in a randomized trial in adults, but currently, there are no similar studies focused on infants [6]. Although synergistic deleterious effects of CMV and TB coinfection have been reported, the impact of these coinfections is poorly understood in children [7–10].

The ongoing EMPIRICAL trial is an international phase II-III multicenter, open-label randomized factorial clinical trial. This trial is funded by EDCTP (Grant number: RIA2017MC- 2013 EMPIRICAL). This trial is being conducted in six African countries, Ivory Coast, Malawi, Mozambique, Uganda, Zambia, and Zimbabwe, in collaboration with research organizations from Spain, France, the UK, Italy, and The Netherlands. Participants in the trial were randomized to four different arms in a 1:1:1:1 fashion (SoC, TB-Treatment + SoC, Valganciclovir + SoC, and TB-Treatment + Valganciclovir + SoC). This trial aimed to assess which of these interventions, or combinations of these interventions, decrease mortality in CLHIV admitted with severe pneumonia [11]. Children with a suspected TB or with a close contact of

TB were excluded from enrolment, because according guidelines, they should receive treatment against TB, and therefore cannot be randomized to an arm without TB-T.

This manuscript reports the details of the prespecified statistical analysis plan (SAP) agreed upon by the independent Data and Safety Monitoring Board (DSMB). The DSMB has assessed the trial every 6 months since study onset and recommended trial continuation on each occasion.

## Methods and design

This SAP corresponds to the version v1.1 31 st October 2022 according to the EMPIRICAL study protocol v2.0 10–11–2020. The first version of the SAP (v1.0) was released on 26 th November 2020 and updated to v1.1 on 31 st October 2022 after the interim analysis report dated on 11 th October 2022. The changes on the original version were made following independent DSMB members suggestions to facilitate results visualizations including summary tables for factorial analysis and reporting number of events in figures.

The SAP will ensure that the analysis is not data driven or selectively reported. The results of the primary analysis are expected to occur in July 2025, after all participants enrolled have completed the 360 days of follow-up.

The EMPIRICAL trial is a 2 × 2 factorial, superiority, unlabeled randomized controlled trial. The 1:1:1:1 randomization was stratified using block sizes by center and severity. Severe cases were considered to have any of the following characteristics at randomization: (i) unable to drink/breastfeed, (ii) persisting vomiting, (iii) having convulsions, (iv) lethargic or unconscious, (v) stridor while calm, (vi) severe malnutrition, (vii) central cyanosis, and (viii) oxygen saturation of less than 90%. A factorial clinical trial was proposed for this clinical trial to examine two or more different interventions in the same trial. Each intervention has a different mechanism of impact on the primary endpoint. In this case, we considered it to be more efficient to run one large factorial trial addressing two questions rather than two separate trials addressing a single trial/question individually. Furthermore, if an unanticipated interaction between the interventions exists, a factorial design allows such interactions to be identified, and the relative contribution of each can be explored [12]. The 2 × 2 factorial design allows two primary comparisons: (i) the effect of valganciclovir on mortality and (ii) the effect of TB-Treatment on mortality.

The study is a phase II-III study as follows:

- The empirical treatment for CMV (valganciclovir) is phase II according to the US Food and Drug Administration (FDA) definitions, as the purpose is to inves-

tigate the efficacy and side effects of this treatment in up to several hundred people with the disease/condition (presumed CMV pneumonia), lasting several months to 2 years.

- The empirical treatment for TB (isoniazid, rifampicin, pyrazinamide, and ethambutol) is phase III according to FDA definitions since the trial tries to demonstrate whether a product offers a treatment benefit to a specific population (in this case, HIV-infected infants with unknown-onset severe pneumonia). Phase III studies typically involve 300 to 3000 participants.

All interventions will include the current WHO SoC guidelines to treat pneumonia in children living with HIV (CLHIV) caused by *Pneumocystis jirovecii*, *Streptococcus pneumoniae*, and *Haemophilus influenzae* b [3]. Concomitant SoC therapy includes antibiotic treatment (ceftriaxone or alternatively ampicillin plus gentamicin), cotrimoxazole, prednisolone, antiretroviral treatment for HIV, and indication of TB treatment on symptom-based or microbiologically confirmed. If the child is naïve, or not taking prescribed ART, ART will be started in all HIV-infected infants according to the WHO and national guidelines on day 15 ± 7. ART regimens will be based on what is being used in national programs.

The inclusion and exclusion criteria as well as the endpoints of the study and clinical definitions were described in the protocol [11].

The principal analysis will be based on *intention-to-treat* (ITT). We will analyze patients in the groups they were randomized to, regardless of treatment received after randomization. A sensitivity analysis will be performed, including a per-protocol analysis considering only data of participants who follow the randomization protocol correctly and excluding those with randomization errors (participants received an IMP different than the one in their allocated arm).

In the case of a new TB diagnosis after randomization, participants will be treated with TB-T according to SoC practices and local protocols. These participants will be considered in their allocation arms in both ITT analysis and per-protocol analysis because they follow their intervention allocation (IMP + SoC or SoC).

### Primary objectives

To compare the impact on 15-day and 1-year mortality of combined systematic empirical treatment against TB and CMV plus SoC versus SoC in HIV-infected infants with severe pneumonia.

### Secondary objectives

Secondary objectives are (i) to compare the cumulative days of oxygen therapy from randomization until

discharge of the intervention arms versus SoC, (ii) to compare the cumulative number of days of hospitalization 1 year after randomization of the intervention arms versus SoC, (iii) to evaluate the safety of empirical valganciclovir and empirical TB-T in HIV-infected infants hospitalized with severe pneumonia of the intervention arms versus SoC, (iv) to know the prevalence of CMV infection in HIV-infected infants hospitalized with severe pneumonia, (v) to know the prevalence and incidence of confirmed and unconfirmed TB in HIV-infected infants hospitalized with severe pneumonia, (vi) to know the prevalence of CMV infection and confirmed and unconfirmed TB in HIV-infected children hospitalized with severe pneumonia that died, (vii) to assess the decrease of the quantitative CMV viral load in blood and saliva in HIV-infected infants hospitalized with severe pneumonia treated with valganciclovir, (viii) to assess the diagnostic accuracy of TB-LAM for the diagnosis of confirmed TB (reference: positive Xpert MTB/RIF Ultra in feces and/or nasopharyngeal aspirate (NPA)), and (ix) to analyze the cost-effectiveness of the proposed treatment strategies in each context. Secondary objectives are further detailed in the EMPIRICAL trial protocol [11].

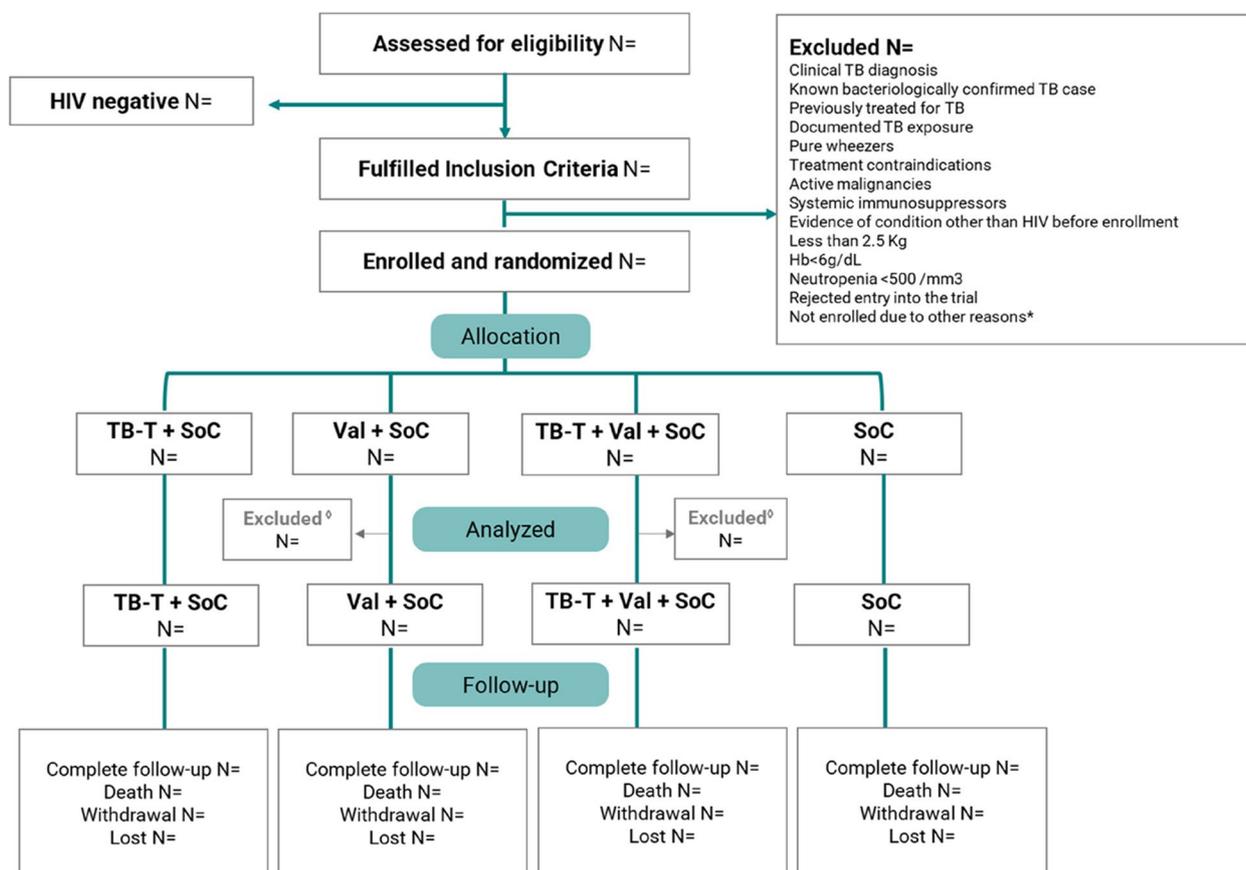
### Hypothesis

Empirical treatment against CMV with oral valganciclovir and empirical TB-Treatment together with standard pneumonia treatment improve survival in HIV-infected infants with severe pneumonia, with a low-risk/benefit profile.

### Primary analysis

To describe the participants included in the study, a flow diagram was constructed according to the Consolidated Standards of Reporting Trials Statement [13]. This flow diagram includes the numbers of participants screened, the numbers eligible, the reasons for ineligibility, and the number of patients randomized and analyzed for the primary outcome, as shown in Fig. 1.

To describe the clinical trial population, the data will be displayed as a whole and summarized by each intervention (arms with/out valganciclovir, arms with/out empirical TB treatment). Baseline demographic characteristics will be included to summarize all subjects. In summary, tables of continuous variables, medians and interquartile ranges will be assessed for nonparametric variables, and means and standard deviations will be presented for parametric variables. The Shapiro–Wilk test will be performed to test normality, and visual methods may also be considered where results from the test will be borderline or inconclusive. In summary, tables of categorical variables, counts and percentages will be used. The denominator for each percentage will be the number of



**Fig. 1** Study flow diagram template for the EMPIRICAL trial (based on the CONSORT flow diagram). HIV, human immunodeficiency virus; TB-T, tuberculosis treatment; Val, valganciclovir; SoC, standard of care; TB, tuberculosis; Hb, hemoglobin

subjects within the population group without considering missing observations unless otherwise specified using the compareGroups R package [14]. Table 1 shows the demographic information by the randomization arm. Table 2 shows the characteristics of the participants who presented the primary endpoint (mortality) during the study.

To analyze the probability of survival among treatment arms, a factorial design was used. Two comparisons will be conducted: (i) participants who were allocated to arms with TB-Treatment (TB-Treatment + SoC and TB-Treatment + Valganciclovir + SoC) compared to those who were not allocated to arms with TB treatment (Valganciclovir + SoC and only SoC) and (ii) participants who were allocated to arms with valganciclovir (Valganciclovir + SoC and TB-Treatment; + Valganciclovir + SoC) compared to those who were not allocated to arms with valganciclovir (TB-Treatment + SoC and only SoC). An interaction between the two main treatments is expected to be synergistic because each treatment focuses on different targets of possible causes of mortality and due to deleterious effect of CMV in the immune response to

TB. Rather than the common assumption in factorial trials of no interaction between the two comparisons, we estimated a better performance in patients who received both treatments. For that reason, both at-the-margins and inside-the-table will be presented as recommended in factorial designs [12] (Tables 3 and 4).

The Kaplan–Meier survival curves will be plotted to compare the probability of survival in each intervention, as presented in Fig. 2. A proportional Cox regression model will be fitted for the primary outcome (mortality). A multivariate flexible parametric survival regression analysis using Weibull distribution will be performed implemented in flexsurv R package [15], including site, CD4 T-cell count, HIV viral load, age at HIV diagnosis, oxygen support use, and nutrition status as possible covariates. The HIV viral load and CD4 percentage will be included in the model as time-dependent covariates using the survival R package [16]. The best final model will be selected using recursive feature elimination via random forest and tenfold cross-validation with 50 repeats implemented in the caret R package [17]. Univariable and multivariable

**Table 1** Baseline characteristics of the participants

Characteristic	All	No valganciclovir	Valganciclovir	Empirical TB-Treatment	No empirical TB-Treatment
	N=	N=	N=	N=	N=
Enrolling site (n, %)					
PACCI (Ivory Coast) (n, %)					
FM-CISM (Mozambique) (n, %)					
UEM (Mozambique) (n, %)					
LSTM-MLW (Malawi) (n, %)					
LMRFT (Malawi) (n, %)					
MU (Uganda) (n, %)					
HerpeZ (Zambia) (n, %)					
UZ-CRC (Zimbabwe) (n, %)					
Severe pneumonia criteria (n, %)					
Nonimprovement with oral treatment (n, %)					
Chest indrawing (n, %)					
Unable to drink/breastfeed (n, %)					
Persisting vomiting (n, %)					
Having convulsions (n, %)					
Lethargic or unconscious (n, %)					
Stridor while calm (n, %)					
Severe malnutrition (n, %)					
Central cyanosis (n, %)					
O2 saturation < 90% (n, %)					
Child age					
Months (median, IQR)					
Gender					
Female (n, %)					
Baseline weight for height					
Z score (median, IQR)					
HIV newly diagnosed at enrollment					
Yes (n, %)					
Oxygen saturation < 90% at arrival					
Yes (n, %)					

**Table 1** (continued)

Characteristic	All	No valganciclovir	Valganciclovir	Empirical TB-Treatment	No empirical TB-Treatment
	N=	N=	N=	N=	N=
Respiratory support					
No (n, %)					
Nasal cannula (n, %)					
CPAP/MV (n, %)					
Age at HIV diagnosis					
Months (median, IQR)					
Virologic and immunologic status					
Viral load at enrollment (Log10)					
HIV viral load at enrollment < 5 logs n (%)					
HIV viral load at enrollment < 5 logs n (%)					
CD4% at enrollment					
ART regimen after enrollment					
None					
2 NRTI + lopinavir/ritonavir					
3 NRTI					
2 NRTI + nevirapine					
2 NRTI + dolutegravir					

PACCI Programme PAC-CI de Côte d'Ivoire, *FM-CISM* Centro de Investigação em Saúde de Manhiça (Mozambique), *UEM* Universidade Eduardo Mondlane Maputo (Mozambique), *LSTM-MLW* Malawi-Liverpool-Wellcome Trust Clinical Research, *LMRFT* Lilongwe Medical Relief Fund Trust (Malawi), *MU* Mulago Hospital (Uganda); *HerpeZ*, Infectious Disease Research & Training in Zambia, *UZ-CRC* University of Zimbabwe Clinical research Centre, *ART* antiretroviral treatment, *SoC* standard of care, *TB-T* tuberculosis treatment, *NRTI* nucleoside reverse-transcriptase inhibitors, *NNRTI* non-nucleoside reverse-transcriptase inhibitors

hazard ratio estimates and 95% confidence intervals will be presented in a forest plot, as shown in Fig. 3. An initial regression model will be fitted, including the interaction and the associated  $p$  value. If there is no evidence of a statistically significant interaction term (i.e.,  $p \geq 0.05$ ), a factorial analysis will be used to determine the success of the trial by analyzing each of the intervention's endpoints independently as two trials. If the interaction term is statistically significant, the effect of each of the four interventions will be tested. The Number Needed to Treat (NNT) will be estimated using the difference in Restricted Mean Survival Time (RMST) between intervention groups, integrating the adjusted survival curves from the Weibull model up to a pre-defined follow-up time. RMST values will be derived separately for each imputed dataset, and the NNT will then be derived by dividing the follow-up time by the RMST difference. Confidence intervals will be obtained

using Rubin's rules to account for variability across imputations.

To analyze snapshot clinically relevant timepoints in the trial (15-day mortality and 1-year mortality), univariable and multivariable logistic regressions will be performed.

To describe the rate, the number of incident deaths (number of deaths) will be calculated in the cohort, and the total follow-up time (person-time) will be computed; death-free individuals in the cohort will be observed over the study period. The estimated mortality incidence will be obtained by dividing the number of deaths by the total duration of follow-up (person-year).

### Secondary analysis

If there is no evidence of a statistically significant interaction effect for the primary outcome, then no interaction effect will be assumed for the secondary outcomes.

**Table 2** Characteristics of participants who died according to randomization arms

	All-case fatality risk/site <i>N</i> = (%)	No valganciclovir <i>N</i> = (%)	Valganciclovir <i>N</i> = (%)	Empirical TB-Treatment <i>N</i> = (%)	No empirical TB-Treatment <i>N</i> = (%)
Enrolling site ( <i>n</i> , %)					
PACCI (Ivory Coast) ( <i>n</i> , %)					
FM-CISM (Mozambique) ( <i>n</i> , %)					
UEM (Mozambique) ( <i>n</i> , %)					
LSTM-MLW (Malawi) ( <i>n</i> , %)					
MU (Uganda) ( <i>n</i> , %)					
HerpeZ (Zambia) ( <i>n</i> , %)					
UZ-CRC (Zimbabwe) ( <i>n</i> , %)					
LMRFT (Malawi) ( <i>n</i> , %)					
Cause ( <i>n</i> , %)					
Cause of death 1					
Cause of death 2					
Cause of death 3					
Cause of death 4					
Cause of death 5					
Time to death (median, IQR)					
Days					
≤ 2 days ( <i>n</i> , %)					
Age at death (median, IQR)					
Months					

*PACCI* Programme PAC-CI de Côte d'Ivoire, *FM-CISM* Centro de Investigação em Saúde de Manhiça (Mozambique), *UEM* Universidade Eduardo Mondlane Maputo (Mozambique), *LSTM-MLW* Malawi-Liverpool-Wellcome Trust Clinical Research, *LMRFT* Lilongwe Medical Relief Fund Trust (Malawi), *MU* Mulago Hospital (Uganda), *HerpeZ* Infectious Disease Research & Training in Zambia, *UZ-CRC*, University of Zimbabwe Clinical research Centre

**Table 3** 2 × 2 factorial analysis

Randomization of TB-Treatment	Randomization of valganciclovir		
	Yes	No	
Yes	<i>Both</i> <i>N</i> =   deaths =	<i>Only TB-Treatment</i> <i>N</i> =   deaths =	<i>All TB-Treatment</i> <i>N</i> =   deaths =
No	<i>Only valganciclovir</i> <i>N</i> =   deaths =	<i>Neither valganciclovir nor TB-Treatment: SoC</i> <i>N</i> =   deaths =	<i>All non-TB-Treatment</i> <i>N</i> =   deaths =
Total	<i>All valganciclovir</i> <i>N</i> =   deaths =	<i>All non-valganciclovir</i> <i>N</i> =   deaths =	Analysis "at the margins"

*TB*, tuberculosis

Likewise, if evidence of a statistically significant interaction effect is identified for the primary outcome, then the secondary outcomes will be analyzed assuming an interaction effect is present.

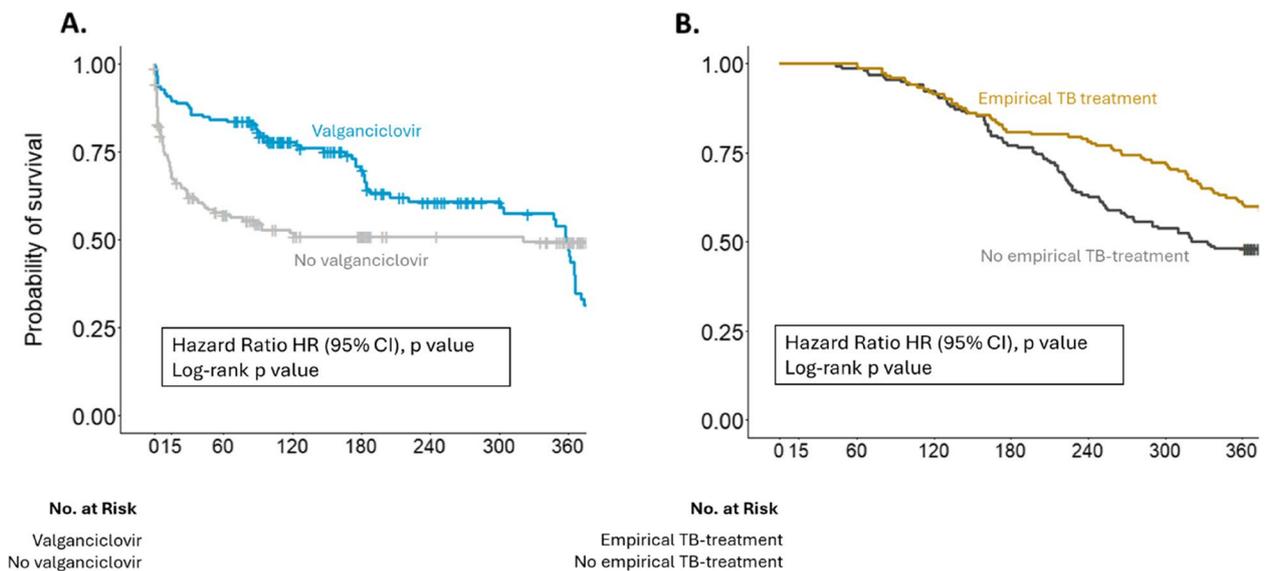
To compare the cumulative days on oxygen therapy from randomization until discharge, a negative binomial model using offset for follow-up time will be used. Similar analyses will be performed to test the differences in cumulative hospitalization days 1 year after randomization.

Serious adverse events (SAEs) will be described among the different interventions in Table 5. The type of registrable SAE, action required, and resolution times will be compared. The chi-square test and Fisher's test (any expected cell count is less than 5) will be performed for categorical variables. For normally distributed continuous variables, Student's *t*-test will be performed, and the Mann-Whitney *U* or Kruskal-Wallis test will be used when nonparametric. All hypothesis testing will be carried out at the 5% significance level and *p* values will be

**Table 4** Summary of the effects presented in the factorial survival analysis

Univariable model		Treatment	
Treatment	Hazard ratio (95% CI)	Treatment	Hazard ratio (95% CI)
1/Both (Val + TB-T) + SoC vs. SoC		3/Only Val + SoC vs. SoC	
2/Only TB-T + SoC vs. SoC		Arms with Val vs. No Val	
Arms with TB-T vs. No TB-T		TB-T + Val + SoC vs. TB-T + SoC	
Both (Val + TB-T) + SoC vs. Val + SoC			
Multivariable model		Treatment	
Treatment	Hazard ratio (95% CI)	Treatment	Hazard ratio (95% CI)
Only TB-T + SoC vs. SoC		Only Val + SoC vs. SoC	
Arms with TB-T vs. No TB-T		Arms with Val vs. No Val	
Only (Val + TB-T) + SoC vs. Val + SoC		TB-T + Val + SoC vs. TB-T + SoC	

TB-T tuberculosis treatment, Val valganciclovir, SoC standard of care, CI confidence interval



**Fig. 2** Kaplan–Meier survival estimates at day 15 and day 360 among the randomization arms

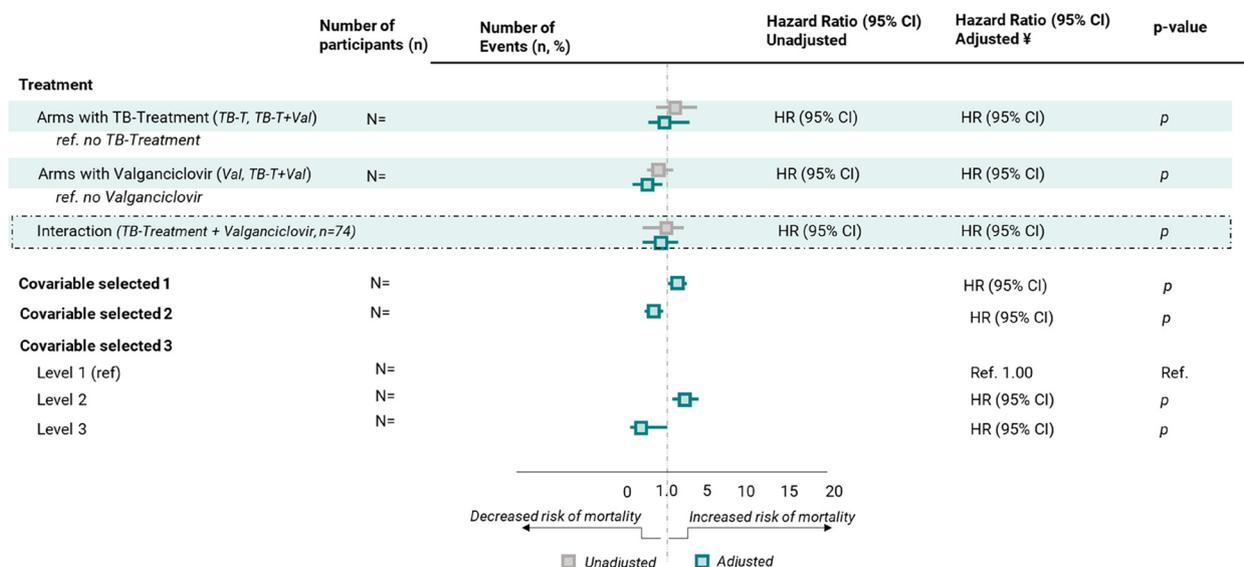
rounded to three decimal places. In summary tables, *p* values less than 0.001 will be reported as <0.001 according to the compareGroups R package [14]. To analyze the association between SAEs and interventions, the person-month density incidence will be calculated using a Poisson regression model. Overdispersion will be checked and if overfitting of zeros in the model, a zero-inflated model will be performed. SAE incidence rate ratios together with 95% CI will be presented. Volcano plots will be used to represent risk differences and significance for the most frequent and significant diseases using the ggplot2 R package [18], as shown in Fig. 4. f

The prevalence of CMV infection will be calculated taking into account 57.1 copies/mL as the cutoff for positivity. CMV-attributable pneumonia will be considered in cases with CMV viral load >4.1 log copies/mL [19]. CMV viral decay will be analyzed using paired Wilcoxon

signed-rank tests. The prevalence and incidence of TB will also be analyzed according to the protocol definitions. The diagnostic accuracy of the TB tests will be assessed using confusion matrices, and the reporting accuracy, sensitivity, specificity, and positive and negative predictive values will be assessed using the caret R package [17]. Results derived from the Xpert Ultra test will be considered as the gold standard.

**Sensitivity analysis**

Sensitivity analysis will be conducted to assess the robustness of the primary trial results by repeating the primary survival regression analysis while taking into account the out-of-randomization scheme, if any. The per-protocol analysis will be performed considering patients who were prescribed an intervention different than it was allocated to. A second sensitivity analysis will be conducted for the



**Fig. 3** Survival estimates (hazard ratios) of the factorial analysis

ITT population to test missing-not-at-random assumptions if more than 20% of the data included in the primary endpoint model are missing.

**Handling of missing data**

To avoid loss of information and statistical power in the association analysis, missing data will be imputed using joint multivariate normal distribution multiple imputation implemented in mice R package [20]. To prevent many assumptions, only variables with less than 20% missing information will be considered for imputation. To obtain a better understanding of the way missing data are distributed among variables in the study, correlation matrices, patching patterns, and box plot analyses will be performed by means of several functions implemented in the MICE and VIM R packages [21].

**Protocol deviations**

Summary of serious protocol deviations will be reported in the biannually DSMB reports and final results of the study.

**Interim analysis**

An interim analysis was planned when 50% of the sample size or 50% of the recruitment time elapsed whatever happen earlier. The symmetric O’Brien-Fleming stopping boundary was applied to the primary endpoint analysis for harm and efficacy.

**Final analysis**

Final analysis is planned to be conducted in July 2025 after database lock on 31 st January 2025.

**Sample size**

The target sample size is 624 randomized participants (156 per arm). Sample size calculations were calculated to reach the main target points of the trial, the short-term (15-day) mortality reduction due to valganciclovir treatment and the long-term (1-year) mortality reduction due to TB-Treatment. The clinical prior assumptions were to observe a reduction from 35% baseline short-term mortality in the SoC to 23% in the valganciclovir-treated patients and a reduction from 41.5% baseline long-term mortality in the SoC to 28.8% in the TB-treated patients, as detailed in the EMPIRICAL trial protocol [22]. The sample size was estimated using the WebPower R package [23] based on 80% statistical power and a 5% two-sided statistical significance level.

**Statistical packages**

All analyses will be carried out using R statistical programming software [24].

**Data management plan**

The study data will be managed according to the data management plan (DMP) of EMPIRICAL study. The study data collection will be recorded and managed using the electronic data capture software of RED-Cap [25] version 8.4.4. Data will be reviewed on regular basis using standardized data quality controls in an effort to increase the cleanliness of future data stored. The quality controls will be conducted for the first five patients on each site and afterward, every 6 months corresponding to each DSMB reporting period. An R code pipeline will be run checking for missing

**Table 5** Description of serious adverse events

	[All]	No valganciclovir	Valganciclovir	Empirical TB-Treatment	No empirical TB-Treatment	<i>p</i>
	<i>N</i> =	<i>N</i> =	<i>N</i> =	<i>N</i> =	<i>N</i> =	
SAE criteria						
Results in death						
Life-threatening						
Requires hospitalization						
Results in persistent or significant disability or incapacity						
Another important medical condition						
Type or registrable SAE						
Any cytopenia						
<i>Neutropenia</i>						
<i>Anemia</i>						
<i>Thrombocytopenia</i>						
<i>Lymphopenia</i>						
<i>Unspecified leukopenia</i>						
Any infectious disease requiring antimicrobial treatment						
<i>New diagnosis TB</i>						
Any liver functions alterations						
Any renal functions alterations						
Any neurologic alterations						
Suspected TB-related IRIS						
Other <sup>a</sup>						
Action taken <sup>b</sup>						
None						
Dose modification						
Medical Intervention						
Hospitalization						
Valganciclovir discontinued						
TB discontinued						
Outcome						
Resolved						
Recovered with minor sequelae						
Recovered with major sequelae						
Continuing treatment						
Condition worsening						
Death						
Unknown						
Pending						
Time resolution (median, IQR)						
Days						
IMP relatedness <sup>¥</sup>						
SAE not related to IMP						
SAE related to IMP (SAR)						
Enrolling site ( <i>n</i> , %)						
PACCI (Ivory Coast) ( <i>n</i> , %)						
FM-CISM (Mozambique) ( <i>n</i> , %)						
UEM (Mozambique) ( <i>n</i> , %)						
LSTM-MLW (Malawi) ( <i>n</i> , %)						
LMRFT (Malawi) ( <i>n</i> , %)						
MU (Uganda) ( <i>n</i> , %)						

**Table 5** (continued)

	[All]	No valganciclovir	Valganciclovir	Empirical TB-Treatment	No empirical TB-Treatment	<i>p</i>
	<i>N</i> =	<i>N</i> =	<i>N</i> =	<i>N</i> =	<i>N</i> =	

HerpeZ (Zambia) (*n*, %)

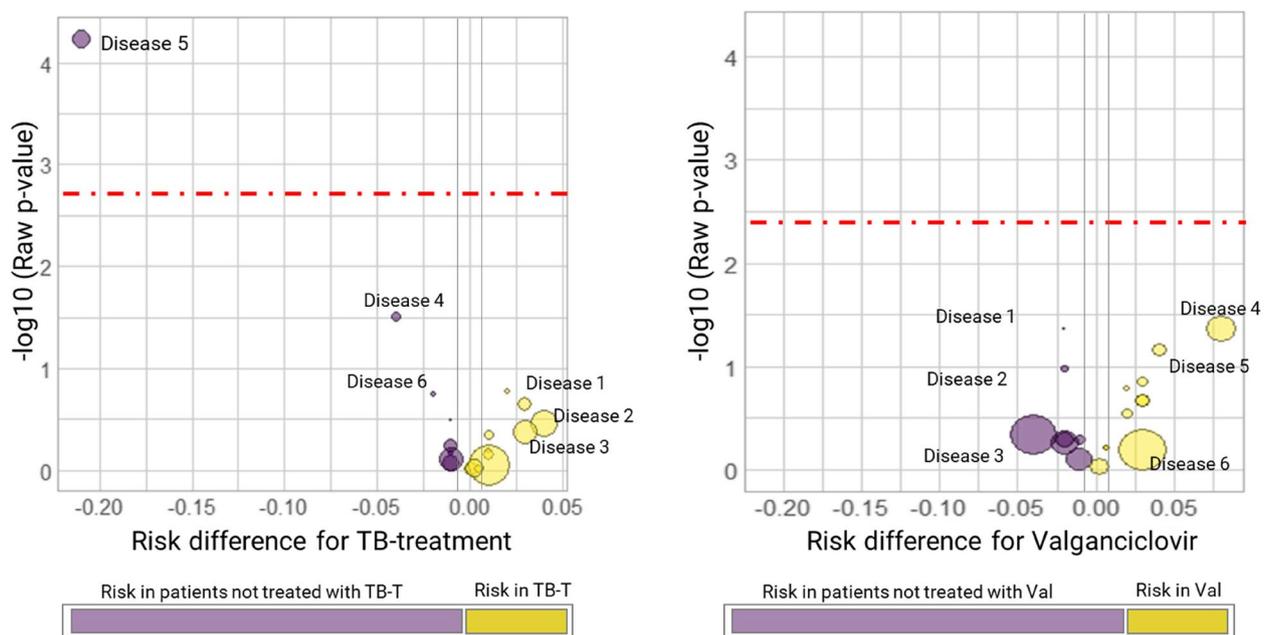
UZ-CRC (Zimbabwe) (*n*, %)

TB-T tuberculosis treatment, Val valganciclovir, SoC standard of care, PACCI Programme PAC-CI de Côte d'Ivoire, FM-CISM Centro de Investigação em Saúde de Manhiça (Mozambique), UEM Universidade Eduardo Mondlane Maputo (Mozambique), LSTM-MLW Malawi-Liverpool-Wellcome Trust Clinical Research, LMRFT Lilongwe Medical Relief Fund Trust (Malawi), MU Mulago Hospital (Uganda), HerpeZ, Infectious Disease Research & Training in Zambia, UZ-CRC University of Zimbabwe Clinical research Centre

<sup>a</sup> Other

<sup>b</sup> One SAE may have several actions associated

¥No *p* value was computed due to unblindness of the study



**Fig. 4** Volcano plot of the serious adverse events risk difference for each treatment

information, clinical inconsistencies, safety inconsistencies, ranges checks, and dates inconsistencies. Queries will be orderly reported in data validation sheets and send it to each hospital/recruiting site responsible for resolving them in REDCap electronic case reports forms.

**Discussion**

In this manuscript, we describe the SAP for the EMPIRICAL clinical trial. The trial is ongoing, and recruitment is already finalized; however, follow-up of the enrolled participants will be completed in January 2025. The publication of the SAP aims to reduce the risk of any potential reporting bias and increase the

transparency of the planned statistical analysis. Any deviations from the current SAP will be justified in the final study report or publications.

**Roles and responsibilities**

Sara Domínguez Rodríguez (Fundación para la Investigación Biomédica del Hospital 12 Octubre and Professor in Universidad Europea de Madrid) is the statistician responsible for writing the SAP. David Lora de Pablos (Fundación para la Investigación Biomédica del Hospital 12 Octubre and Professor in Universidad Complutense de Madrid) is the senior statistician responsible for the SAP. Pablo Rojo Conejo (Hospital Universitario 12 Octubre—Servicio Madrileño de Salud

and Professor in Universidad Complutense de Madrid) is the chief investigator of the study.

### Trial status

The trial finished the recruitment period on 31 st January 2024 but data collection is still active for participants' follow-up information. The protocol version linked to this statistical analysis plan is version 2.0 10 th November 2020. The authors published the study protocol and SAP in different independent manuscripts to provide more detail on statistical analysis due to the factorial design complexity.

### Authors' contributions

SDR and DL elaborate the statistical design; AB data management and tables specifications; AT, CT, LM, and PR protocol writing; OM, VL, AN, DB, QB, MB, RM, PIT, TM, JM, WCB, JS, VM, CC, HAM principal investigators of the study, participate and review the study protocol and review the manuscript.

### Funding

This study has been funded by EDCTP (Grant number: RIA2017MC- 2013 EMPIRICAL). The funder will have no role in the trial design, conduct, data analysis and interpretation, manuscript writing, and dissemination of results.

### Data availability

CTU will oversee the intra-study data sharing process, and they will be given access to the cleaned data sets. Project data sets will be housed in the CTU secure server allocated in Hospital 12 Octubre (Madrid, Spain), and all data sets will be encrypted and password protected. Project PIs will have direct access to their own site's data sets and will have access to other sites' data by justified request. To ensure confidentiality, data dispersed to study staff will be blinded to any identifying participant information. REDCap software implements an audit trail to ensure the data access tracking by the study personnel. The scientific integrity of the project requires that the data from all the sites be analyzed study-wide and reported as such. Thus, an individual center is not expected to report the data collected from its center alone. All presentations and publications are expected to protect the integrity of the major objective(s) of the study. No later than 3 years after the collection of the 1-year post-randomization visits, the CTU will deliver a pseudo-anonymized data set and metadata to an appropriate data archive for sharing purposes unless specific national legislation from any of the sites impedes sharing open access of the data. In this case, the dataset of this site will not be released.

### Declarations

#### Ethics approval and consent to participate

The trial has been approved by the Investigation Ethics Committee of Medicines (CEIm) N°CEIm: 19/096 Hospital 12 de Octubre on March 13, 2019, and by all the enrolling sites' ethical boards and regulatory agencies.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Instituto de Investigación Sanitaria Hospital (imas12), Fundación Biomedica del Hospital Universitario (FIB-H120), Madrid, Spain. <sup>2</sup>Universidad Europea, Madrid, Spain. <sup>3</sup>Spanish Clinical Research Network (SCReN), Madrid, Spain. <sup>4</sup>Facultad de Estudios Estadísticos, Universidad Complutense de Madrid (UCM), Madrid, Spain. <sup>5</sup>Servicio de Pediatría, Hospital Universitario Infanta Sofía, Servicio Madrileño de Salud (SERMAS), Madrid, Spain. <sup>6</sup>Servicio de Pediatría, Hospital Universitario, Servicio Madrileño de Salud (SERMAS), 12 de Octubre, Madrid, Spain. <sup>7</sup>Department of Infectious Disease Epidemiology, London,

School of Hygiene & Tropical Medicine, London, UK. <sup>8</sup>Université de Bordeaux, Bordeaux Population Health, GHiGS, Inserm U1219, IRD EMR271 Bordeaux, France. <sup>9</sup>Institut National de La Santé Et de La Recherche Médicale (Inserm), CERPOP, University Toulouse 3, Toulouse, France. <sup>10</sup>Penta-Onlus Foundation (PENTA), Padua, Italy. <sup>11</sup>Stichting Katholieke Universiteit—Radboudumc (RUMC), Nijmegen, The Netherlands. <sup>12</sup>ISGlobal, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain. <sup>13</sup>Pediatrics Department, Hospital Sant Joan de Déu, I, Universitat de Barcelona, Barcelona, Spain. <sup>14</sup>ICREA, Pg. Lluís Companys 23, Barcelona 08010, Spain. <sup>15</sup>Centro de Investigação Em Saúde de Manhiça (CISM), Maputo, Mozambique. <sup>16</sup>Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain. <sup>17</sup>University of Lincoln, Lincoln, UK. <sup>18</sup>Unité Pédagogique de Dermatologie Et Infectiologie, UFR Sciences Médicales, Programme PAC-CI, Ivory Coast, Abidjan, Côte d'Ivoire. <sup>19</sup>Kamuzu University Health Sciences, Blantyre, Malawi. <sup>20</sup>Liverpool School of Tropical Medicine (LSTM), Malawi-Liverpool Wellcome Programme (MLW), Blantyre, Malawi. <sup>21</sup>Lilongwe Medical Relief Trust (LMRFT), UNC Project Malawi, Lilongwe, Malawi. <sup>22</sup>David Geffen School of Medicine, University of California Los Angeles, Los Angeles, USA. <sup>23</sup>Universidade Eduardo Mondlane (UEM), Maputo, Mozambique. <sup>24</sup>Department of Paediatrics and Child Health, College of Health Sciences, Makerere University, Kampala, Uganda. <sup>25</sup>Joint Clinical Research Centre, Kampala, Uganda. <sup>26</sup>School of Medicine, University of Zambia, Lusaka, Zambia. <sup>27</sup>University of Zimbabwe, Harare, Zimbabwe. <sup>28</sup>Facultad de Medicina, Universidad Complutense, Madrid, Spain.

Received: 10 May 2024 Accepted: 14 April 2025

Published online: 30 April 2025

### References

- Theodoratou E, McAllister DA, Reed C, Adeloye DO, Rudan I, Muhe LM, et al. Global, regional, and national estimates of pneumonia burden in HIV-infected children in 2010: a meta-analysis and modelling study. *Lancet Infect Dis*. 2014;14(12):1250–8.
- Njuguna IN, Cranmer LM, Otieno VO, Mugo C, Okinyi HM, Benki-Nugent S, et al. Urgent versus post-stabilisation antiretroviral treatment in hospitalised HIV-infected children in Kenya (PUSH): a randomised controlled trial. *Lancet HIV*. 2018;5(1):e12–22.
- Organization WH. Revised WHO classification and treatment of pneumonia in children at health facilities: evidence summaries. 2014;
- Bates M, Shibemba A, Mudenda V, Chimoga C, Tembo J, Kabwe M, et al. Burden of respiratory tract infections at post mortem in Zambian children. *BMC Med*. 2016;14(1):99.
- Thimbleby H. Guidelines for managing advanced HIV disease and rapid initiation of antiretroviral therapy. *Behaviour and Information Technology*. 2017;2(2):2–56.
- Hosseiniour MC, Bisson GP, Miyahara S, Sun X, Moses A, Riviere C, et al. Empirical tuberculosis therapy versus isoniazid in adult outpatients with advanced HIV initiating antiretroviral therapy (REMEMBER): a multicountry open-label randomised controlled trial. *The Lancet*. 2016;387(10024):1198–209.
- Fletcher HA, Snowden MA, Landry B, Rida W, Satti I, Harris SA, et al. T-cell activation is an immune correlate of risk in BCG vaccinated infants. *Nat Commun*. 2016;7(1):1–11.
- Fodil-Cornu N, Vidal SM. Type I interferon response to cytomegalovirus infection: the kick-start. *Cell Host Microbe*. 2008;3(2):59–61.
- O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. *Annu Rev Immunol*. 2013;31:475–527. <https://doi.org/10.1146/annurev-immunol-032712-095939>.
- Nagu T, Aboud S, Rao M, Matee M, Axelsson R, Valentini D, et al. Strong anti-Epstein Barr virus (EBV) or cytomegalovirus (CMV) cellular immune responses predict survival and a favourable response to anti-tuberculosis therapy. *Int J Infect Dis*. 2017;56:136–9.
- Rojo P, Moraleda C, Tagarro A, Domínguez-Rodríguez S, Castillo LM, Tato LMP, et al. Empirical treatment against cytomegalovirus and tuberculosis in HIV-infected infants with severe pneumonia: study protocol for a multicenter, open-label randomized controlled clinical trial. *Trials*. 2022;23(1):1–36.
- Kahan BC. Bias in randomised factorial trials. *Stat Med*. 2013;32(26):4540–9.

13. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340(7748):698–702.
14. Subirana I, Vila J, Sanz H, Gavin L, Penafiel J, Gimenez D. Package *compar*. 2018.
15. Jackson C. *flexsurv*: a platform for parametric survival modeling in R. *J Stat Softw*. 2016;70(8):1–33. <http://www.jstatsoft.org/v70/i08/>. Cited 2 Jan 2019.
16. Kassambara A, Kosinski M. *Survminer*: drawing survival curves using "ggplot2." 2018.
17. Kuhn M. Building predictive models in R using the *caret* package. *J Stat Softw*. 2008;28(5):1–26.
18. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer. 2016.
19. Hsiao NY, Zampoli M, Morrow B, Zar HJ, Hardie D. Cytomegalovirus viraemia in HIV exposed and infected infants: prevalence and clinical utility for diagnosing CMV pneumonia. *J Clin Virol*. 2013;58(1):74–8.
20. van Buuren S, Groothuis-Oudshoorn K. *mice*: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1–67. Available from: <https://www.jstatsoft.org/index.php/jss/article/view/v045i03>. Cited 12 Apr 2024.
21. van Buuren S, Groothuis-Oudshoorn K. *mice*: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45(3):1–67.
22. Rojo P, Moraleda C, Tagarro A, Domínguez-Rodríguez S, Castillo LM, Tato LMP, et al. Empirical treatment against cytomegalovirus and tuberculosis in HIV-infected infants with severe pneumonia: study protocol for a multicenter, open-label randomized controlled clinical trial. *Trials*. 2022;23(1):531.
23. Zhang Z, Mai Y. *WebPower*: basic and advanced statistical power analysis. 2018.
24. R Development Core Team. *A language and environment for statistical computing*. Vienna, Austria; 2008.
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.