Supplementary method details

Study setting

The definition of seasonal and perennial transmission sites was done by the clinical trial team through preliminary assessment of malaria prevalence and transmission level (1). Participants recruited had different vaccination regimes with respect to malaria transmission settings (1). At seasonal transmission sites, children received primary series doses before peak malaria season, ensuring high antibody titre during the peak season. While at perennial transmission sites, participants were administered vaccines with age-based regime during any period of the year (1).

Data cleaning – Definition of antibiotics, selection of study population, definition of observation window

Individual-level data on medication prescriptions were collected by clinicians at trial site clinics through routine follow-up visits, clinic visit records and retrospective self-report. All recorded medications were classified as either antibiotics or non-antibiotics, and antibiotic drugs were further categorised into specific antibiotic classes based on their mechanism of action and chemical structure according to standard guidelines (Table.S7) (2,3). The antibiotic prescription data were assumed to be the only source of participants' antibiotic usage.

This study restricted the analysis to modified-per-protocol population that received the full three primary series doses and the first booster dose with correct intervals (n=4369) (4). This was done initially in Excel, then checked and further restricted in R to ensure that all participants in the analysis have received 3 primary series doses and 1 booster dose, and

meet the modified-per-protocol definition that the booster dose is administered 365 +/- 35 days after third dose (4). Individuals who met the criteria and did not have medication records were assumed as not being prescribed antibiotics during the observation window.

The end date of observation window used the follow-up visit date of the second booster administration or 365 days after the first booster dose vaccination (Table.S1). For participants that had withdrawn from the study due to loss to follow up, withdrawal and death, study exit dates were used as the end date of observation window. The use of follow-up visit dates as the end of observation window followed the priority order of using 2B0, B365, B1Y, study exit date (Table.S1). One participant did not have any recorded follow-up visit dates because the scheduled 2B0 pre-vaccination visit was not conducted due to ongoing serious adverse events. For this participant, the end date was calculated as 365 days after first booster.

Age is categorised into three age groups based on biological development: below 1 year old, 1 to 2 years old, and 2 to 3 years old. This transformation was conducted to account for potential non-linear association of age and antibiotic usage, rather than treating age as a continuous variable. During participant enrollment stage of the clinical trial, the identical age group separation was used and equal number of participants was recruited in each age group (1,4).

Multivariable mixed-effect logistic regression model

This model estimated the odds ratio for antibiotic usage between R21/Matrix-M vaccine recipients and controls, adjusted for age, sex and site-level random effects. Vaccination status was included as the primary explanatory variable, and the outcome variable was defined as a binary indicator of whether an individual received at least one antibiotic

prescriptions during the follow-up period. Akaike Information Criterion (AIC) values and model diagnostics were used to evaluate model fit and compare models without random effects (equation 1), with random intercept only (equation 2), and with both random intercept and various random slopes (equation 3-5), to select the best-fitting model. The goodness-of-fit of the model were assessed by examining and plotting the predicted probabilities from the model against the observed outcomes. If the distribution was a clear positive trend with higher predicted probabilities associated with receiving antibiotics, the model was considered well calibrated.

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \cdot vaccination\ status_{ij} + \beta_2 \cdot age\ group_{ij} + \beta_3 \cdot sex_{ij} + \epsilon_{ij} \quad (1)$$

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \cdot vaccination\ status_{ij} + \beta_2 \cdot age\ group_{ij} + \beta_3 \cdot sex_{ij} + u_{0j} + \epsilon_{ij} \quad (2)$$

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \cdot vaccination\ status_{ij} + \beta_2 \cdot age\ group_{ij} + \beta_3 \cdot sex_{ij} + u_{0j} + u_{1j} \cdot age\ group_{ij} + \epsilon_{ij} \quad (3)$$

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \cdot vaccination\ status_{ij} + \beta_2 \cdot age\ group_{ij} + \beta_3 \cdot sex_{ij} + u_{0j} + u_{2j} \cdot sex_{ij} + \epsilon_{ij} \quad (4)$$

$$\text{logit}(p_{ij}) = \beta_0 + \beta_1 \cdot vaccination\ status_{ij} + \beta_2 \cdot age\ group_{ij} + \beta_3 \cdot sex_{ij} + u_{0j} + u_{1j} \cdot age\ group_{ij} + u_{2j} \cdot sex_{ij} + \epsilon_{ij} \quad (5)$$

Multivariable mixed-effect Poisson or negative binomial regression model

This model estimated the incidence rate ratio of antibiotic prescriptions between R21/Matrix-M vaccinated and control vaccinated groups, adjusted for age, sex and site-level random effects. Vaccination status was included as the primary explanatory variable, and the outcome variable was defined as the total number of antibiotic prescriptions an individual received offsetting by the individual's total duration of follow-up period. The Poisson model was checked for overdispersion, which is the situation when variance of the outcome variable is much higher than mean (5). Overdispersion leads to underestimation of standard errors and incorrect interpretation of the relatioonship (5). Dispersion statistic, which is the ratio between Pearson Chi-square value and residual degrees of freedom was calculated (5). Dispersion statistic value above 1.5 was considered as overdispersion (5). To further validate the detection of overdispersion, Score test (equation 6; Table.1) and Lagrange multiplier test

(equation 7; Table.1) were performed (5). Both tests had the null hypothesis that no dispersion was detected in the model. When *p*-value was less than 0.05, the null hypothesis was rejected. The equations are:

$$z = \frac{(y-\mu)^2 - y}{\mu\sqrt{2}} \tag{6}$$

$$x^2 = \frac{(\sum_{i=1}^{n} \mu_i^2 - n\bar{y}_i)^2}{2\sum_{i=1}^{n} \mu_i^2} \tag{7}$$

Table.1 Equation parameters of Score test and Lagrange multiplier test

Parameters	Definition
z	Overdispersion z score
y	Observed number of antibiotic prescriptions in the data
μ	Predicted number of antibiotic prescriptions under Poisson model
x^2	Chi-square value
n	Number of observations

Similar to the logistic model, AIC values and model diagnostics were used to evaluate model fit and compare models without random effects, with random intercepts only, and with both random intercept and various random slopes, to select the best-fitting model. The predicted number of antibiotic prescriptions were plotted against observed values to evaluate the goodness-of-fit. If the distribution was a positive linear trend with high predicted values associated with high observed values, the model was considered well calibrated.

Survival analysis and Cox regression models

The survival times were defined as: baseline time 0 was the administration date of the third dose in the primary series vaccine, and the end time was set as the end of observation window, which is the time prior to the second booster administration. Time to first antibiotic prescription refers to the time interval between time 0 and the date of the first antibiotic prescription. Since antibiotics prescribed before 14 days after third vaccine dose

administration were excluded, the event is 0 from time 0 to time 14. Participants that did not receive antibiotics prescription during the observation period were censored, with the time of event being the end of observation window. For the recurrent event survival analysis, Andersen-Gill Cox model treated recurrent antibiotic prescriptions as independent events with same baseline hazards and same covariate effects regardless of event orders (6–8). The Cox models assumed that the ratio of the hazards maintained constant and proportional over time (9). Therneau-Grambsch test were performed to check whether the assumptions were violated. The null hypothesis was defined as proportional hazards assumptions being true (9). When *p*-value was above 0.05, the null hypothesis was accepted. Schoenfeld residual analysis was conducted to examine the coefficient and slope line of the hazard ratio. If the coefficient was zero, the slope was a horizontal line and Schoenfeld residuals were randomly scattered around zero over time, the proportional hazards assumptions were considered reasonable and not being violated (9).

Validation of vaccine efficacy

All statistical analyses including descriptive statistics, regression models and survival analyses were reproduced and performed for the association between R21/Matrix-M vaccination and malaria treatment to validate the consistency of this medication prescription data with previously reported vaccine efficacy against clinical malaria. Malaria treatment was defined as any medications prescribed with diagnosis indications of malaria during observation period, including both antibiotics and non-antibiotics.

Technology

Statistical analyses were all performed using R Version 4.4.2 and RStudio Version 2025.05.0+496, with a significance level set at 5% and confidence intervals at 95%.

R packages used in the modelling and analyses included: pacman v0.5.1, readxl v1.4.3, writexl v1.5.4, tidyverse v2.0.0, dplyr v1.1.4, lubridate v1.9.3, ggplot2 v3.5.1, ggpubr v0.6.0, lme4 v1.1-35.5, broom v1.0.7, broom.helpers v1.21.0, broom.mixed v0.2.9.6, gtsummary v2.2.0, forestplot v3.1.6, sjPlot v2.8.17, survival v3.7-0, survminer v0.5.0.

Ethics

The data of this study was from the clinical trial NCT04704830 registered on ClinicalTrials.gov, approved by ethics and regulatory authorities at local sites and Oxford Tropical Research Ethics Committee (1). The trial was conducted in accordance with Medicine for Human use (clinical trials) Regulations 2004, European Communities (Clinical Trial on Medicinal Products for Human Use) Regulations 2004, and principles of Declaration of Helsinki 2013 (1). Since the clinical trial has been extended for ongoing research, to maintain double-blinded, external statisticians removed patient ID, assigned random number to participants when revealing vaccination status.

References

1. Datoo MS, Ewer KJ, Hil AVS, Lopez FR. VAC078 Clinical Trial Protocol: A Phase III randomized controlled multi- centre trial to evaluate the efficacy of the R21/Matrix-M vaccine in African children against clinical malaria. 2025.
2. MSD Manual [Internet]. [cited 2025 May 4]. Overview of Antibiotics. Available from: https://www.msdmanuals.com/home/infections/antibiotics/overview-of-antibiotics
3. REVIVE [Internet]. [cited 2025 May 4]. Classes of Antibiotics. Available from: https://revive.gardp.org/resource/classes-of-antibiotics/?cf=encyclopaedia
4. Datoo MS, Dicko A, Tinto H, Ouédraogo JB, Hamaluba M, Olotu A, et al. Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial. The Lancet. 2024 Feb;403(10426):533–44.
5. Serra I. R Pubs by RStudio. [cited 2025 May 20]. Testing Overdispersion. Available from: https://rpubs.com/DonArres/TestingOverdispersion

6. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. International Journal of Epidemiology. 2015 Feb;44(1):324–33.

7. Ozga AK, Kieser M, Rauch G. A systematic comparison of recurrent event models for application to composite endpoints. BMC Med Res Methodol. 2018 Dec;18(1):2.

8. Yang W, Jepson C, Xie D, Roy JA, Shou H, Hsu JY, et al. Statistical Methods for Recurrent Event Analysis in Cohort Studies of CKD. CJASN. 2017 Dec;12(12):2066–73.

9. Keele L. Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models. Polit anal. 2010;18(2):189–205.