Gates Foundation

Cessation Thresholds for MDA Azithromycin

A secondary analysis of REACH trial data

William Msemburi

August, 2025

1 Introduction

Child mortality remains stubbornly high in sub-Saharan Africa—about 69 deaths per 1,000 live births in 2023 [1] - fourteen times higher than the estimate for Europe and North America and about double the global average. This has driven the search for simple, scalable interventions that can reduce deaths where health systems are weak.

Mass drug administration (MDA) of azithromycin has emerged as a promising approach. Originally used for trachoma elimination, azithromycin MDA showed unexpected benefits in reducing all-cause child mortality [2]. The landmark MORDOR trial found a 13.5% reduction in under-five mortality with biannual treatment across Niger, Malawi, and Tanzania - the first randomized evidence that antibiotics could reduce all-cause child mortality outside of pneumonia treatment. Benefits persisted up to two years after the final treatment round [3].

But the benefits were not consistent across all trials. Impacts were strongest in high-mortality settings, raising questions about effectiveness in lower-risk populations [4]. Meanwhile, concerns about antimicrobial resistance have grown, with evidence of increased macrolide resistance following repeated azithromycin exposure [5].

The WHO tried to balance these concerns in 2020, recommending targeted use only where infant mortality exceeds 60 per 1,000 or under-five mortality exceeds 80 per 1,000 live births [6]. However, these thresholds weren't derived from trial data—they reflected expert judgment attempting to weigh benefits against resistance risks. While expert-based thresholds have precedent in global health programs, empirical approaches using trial data to define cessation criteria have proven more robust and implementable for mass drug administration programs [7, 8, 9].

The aim of our research is simple yet significant: to provide the first data-driven basis for defining cessation criteria in azithromycin mass drug administration (MDA) programs.

Building on established approaches for spatial mortality estimation [10, 11, 12], we conduct a comprehensive secondary analysis of cluster-randomized trial data. Our four-stage Bayesian framework (i) establishes age-specfic mortality relationships from demographic surveys to impute neonate mortality, (ii) derives trial-specific overal mortality estimates, (iii) produces spatially coherent community-level baseline mortality surfaces, and (iv) develops threshold models to identify the mortality levels at which treatment benefits are no longer observed.

Recognizing that mortality alone may not capture all sources of variation, we also extend the analysis to assess whether other contextual factors—such as vaccination coverage, malaria burden, and healthcare access—modify treatment effects. This allows us to situate mortality-based thresholds within the broader landscape of contextual heterogeneity.

Key messages

- **Purpose:** Express cessation thresholds for azithromycin MDA in standard demographic terms (IMR, U5MR) with uncertainty.

- **Key finding on context:** Baseline mortality explains most variation in benefit; other factors correlate with mortality but don't consistently modify treatment effects.

- **Cessation thresholds estimated:** U5MR *mean (95% CrI)* = 49.4 (35.8-72.0) per 1,000; IMR *mean (95% CrI)* = 27.3 (20.3-37.9) per 1,000.

- **Program guidance:** Cease MDA when baseline mortality falls below these thresholds, with antimicrobial resistance surveillance in place.

2   Methods

Our analytical approach addressed four interconnected challenges: establishing reliable age-specific mortality patterns for trial settings, anchoring these patterns in observed trial data, accounting for spatial correlation in community-level mortality, and identifying mortality thresholds where azithromycin benefits disappear. We developed a sequential four-stage Bayesian framework to address these challenges systematically.

2.1   Stage 1: Demographic age patterns from household surveys

Trial clusters typically observe few deaths, making direct estimation of age-specific mortality patterns unstable. To address this, we leveraged complete birth histories from Demographic and Health Surveys (DHS) conducted in trial countries between 2010–2018: Burkina Faso, Malawi, Niger, Nigeria, and Tanzania. These nationally representative surveys provide robust age-specific mortality estimates that can inform trial-based modeling [11, 12].

2.1.1   Age-specific hazard estimation

For each child in the birth histories, we calculated person-time contributions across standard demographic age intervals: neonatal (0 months), early post-neonatal (1–11 months), late post-neonatal (12–23 months), and subsequent yearly intervals through 59 months. Within each age group a, we estimated monthly death probabilities p_a using survey weights to ensure population representativeness.

We converted these probabilities to cumulative hazards, which provide a more natural scale for modeling rare events [13]:

$$H_a = -n_a \log(1 - p_a), \tag{1}$$

where:

H_a = cumulative hazard for age group a

n_a = number of months in age group a

p_a = estimated monthly probability of death in age group a

2.1.2 Neonatal-to-postneonatal bridge

Trial follow-up typically begins at age 1 month, necessitating prediction of neonatal mortality from observed post-neonatal patterns. We established this relationship using pooled DHS data through a log-linear regression model:

$$\log H_{0,c} = \gamma_c + \beta_1 \log H_{1-11,c} + \beta_2 \log H_{12-59,c} + \varepsilon_c, \tag{2}$$

where:

$H_{0,c}$ = neonatal cumulative hazard in country c

$H_{1-11,c}$ = early post-neonatal hazard (months 1–11) in country c

$H_{12-59,c}$ = late post-neonatal hazard (months 12–59) in country c

γ_c = country-specific intercept

β_1, β_2 = regression coefficients (pooled across countries)

ε_c = country-specific residual with estimated variance σ_ε^2

This model provides the foundation for predicting neonatal mortality in trial clusters from their observed post-neonatal patterns, with appropriate uncertainty propagation.

2.2 Stage 2: Trial-anchored baseline priors

While DHS surveys provide robust national age patterns, trial sites may differ systematically from national averages due to selection criteria, geographic clustering, or temporal changes. To ensure our mortality estimates reflect actual trial settings, we calibrated demographic priors using observed placebo data from each trial-country combination [4].

2.2.1 Baseline hazard estimation

For each trial-country group g, we estimated baseline mortality using observed deaths and person-time in placebo clusters. We fit a quasi-Poisson generalized linear model with cluster-robust standard errors to account for overdispersion:

$$\log E[y_{c,g}] = \alpha_g + \log(E_{c,g}), \tag{3}$$

where:

$y_{c,g}$ = observed deaths in placebo cluster c within trial-country group g

$E_{c,g}$ = person-time at risk in cluster c, group g (exposure offset)

α_g = log-hazard intercept for trial-country group g

We converted group-level intercepts to cumulative hazards for the 1–59 month age range:

$$\log H_{g,1-59} = \alpha_g + \log(59/12), \tag{4}$$

where the scaling factor 59/12 converts monthly rates to cumulative hazards over 59 months.

2.2.2 Age decomposition through Beta-binomial updating

To decompose the 1–59 month hazard into early (1–11) and late (12–59) components, we used observed age-specific deaths to update DHS-derived Beta priors. Let p_g denote the fraction of total 1–59 month hazard occurring in months 1–11. With observed deaths $d_{g,1-11}$ and $d_{g,12-59}$ in each age group:

$$p_g \mid \text{data} \sim \text{Beta}(\alpha_{g,0} + d_{g,1-11}, \beta_{g,0} + d_{g,12-59}), \tag{5}$$

where:

p_g = fraction of 1–59 month hazard in early post-neonatal period

$\alpha_{g,0}, \beta_{g,0}$ = DHS-derived Beta prior parameters for trial-country g

$d_{g,1-11}, d_{g,12-59}$ = observed placebo deaths by age group

This Bayesian updating provides trial-calibrated age shares while maintaining connection to established demographic patterns, enabling robust uncertainty quantification throughout subsequent modeling stages.

2.3 Stage 3: Bayesian spatial mortality estimation

Individual trial clusters typically observe few deaths, making direct mortality estimation highly unstable. However, nearby clusters often share similar epidemiological conditions, suggesting that spatial borrowing of information can improve estimation precision while preserving genuine local variation [10, 14]. We developed a hierarchical spatial model that balances these competing objectives through principled regularization.

2.3.1 Hierarchical likelihood structure

For each cluster i in trial-country group g, we modeled observed placebo deaths as:

$$y_i \sim \text{Poisson}(\mu_i), \tag{6}$$

with log-linear predictor:

$$\log \mu_i = \log E_i + \underbrace{[\mu_g + u_i + v_i]}_{\log \lambda_{i,1\text{--}59}}, \tag{7}$$

where:

y_i = observed deaths in placebo cluster i

E_i = person-time at risk in cluster i (exposure offset)

μ_g = trial-country group mean (from Stage 2 calibration)

u_i = smooth spatial deviation for cluster i

v_i = sparse local deviation for cluster i

$\lambda_{i,1\text{--}59}$ = cluster-specific mortality rate (ages 1–59 months)

2.3.2 Spatial process specification

The smooth spatial component u_i follows a Matérn covariance process with smoothness parameter $\nu = 3/2$, implemented through Hilbert Space Gaussian Process (HSGP) approximation [15, 16]. This approach provides computational efficiency while maintaining flexibility in capturing geographic mortality patterns.

The spatial process is defined by:

$$u(\mathbf{s}) \sim \text{GP}(0, C(\mathbf{s}, \mathbf{s}'; \rho, \sigma_u^2)) \tag{8}$$

$$C(\mathbf{s}, \mathbf{s}'; \rho, \sigma_u^2) = \sigma_u^2 \left(1 + \frac{\sqrt{3}\|\mathbf{s} - \mathbf{s}'\|}{\rho} \right) \exp\left(-\frac{\sqrt{3}\|\mathbf{s} - \mathbf{s}'\|}{\rho} \right), \tag{9}$$

where:

$\mathbf{s}$ = spatial coordinates (longitude, latitude)

ρ = spatial range parameter

σ_u^2 = spatial variance parameter

$C(\cdot, \cdot; \rho, \sigma_u^2)$ = Matérn 3/2 covariance function

2.3.3 Local deviation regularization

The local deviation terms v_i use regularized horseshoe priors [17, 18] to distinguish genuine outliers from noise:

$$v_i \sim \text{Normal}(0, \tau \tilde{\lambda}_i) \tag{10}$$
$$\tilde{\lambda}_i \sim \text{CauchyPos}(0, 1) \tag{11}$$
$$\tau \sim \text{CauchyPos}(0, \tau_0), \tag{12}$$

where:

τ = global shrinkage parameter

$\tilde{\lambda}_i$ = local shrinkage parameter for cluster i

τ_0 = prior scale based on expected sparsity

CauchyPos$(0, \sigma)$ = positive half-Cauchy distribution with scale σ

This specification provides adaptive shrinkage: clusters with weak evidence for local departures are shrunk toward the spatial mean, while genuine outliers are preserved.

2.3.4 Age decomposition

To recover standard demographic indicators (IMR, U5MR), we decomposed the fitted 1–59 month hazard into constituent age groups. Using the trial-calibrated Beta-binomial shares from Stage 2:

$$p_g \mid \text{data} \sim \text{Beta}(\alpha_{g,0} + d_{g,1-11}, \beta_{g,0} + d_{g,12-59}), \tag{13}$$

we computed age-specific rates:

$$\log \lambda_{i,1-11} = \log \lambda_{i,1-59} + \log p_g + \log(59/11) \tag{14}$$
$$\log \lambda_{i,12-59} = \log \lambda_{i,1-59} + \log(1 - p_g) + \log(59/48), \tag{15}$$

where the scaling factors ensure proper hazard partitioning. Converting to cumulative hazards:

$$H_{i,1-11} = \lambda_{i,1-11} \times (11/12) \tag{16}$$
$$H_{i,12-59} = \lambda_{i,12-59} \times (48/12) \tag{17}$$
$$H_{i,1-59} = H_{i,1-11} + H_{i,12-59} \tag{18}$$

2.3.5   Neonatal prediction

Finally, we predicted neonatal hazards using the country-specific bridge relationships from Stage 1:

$$\log H_{i,0} = \alpha_c + \beta_1 \log H_{i,1-11} + \beta_2 \log H_{i,12-59} + \varepsilon_c, \tag{19}$$

where $(\alpha_c, \beta_1, \beta_2, \sigma_\varepsilon)$ are fixed at DHS-derived values with appropriate uncertainty propagation.

The complete model yields posterior distributions over cluster-specific hazards $\{H_{i,0}, H_{i,1-11}, H_{i,12-59}\}$, enabling calculation of standard demographic indicators with full uncertainty quantification:

$$\text{IMR}_i = 1000 \times (1 - \exp(-H_{i,0})) \tag{20}$$
$$\text{U5MR}_i = 1000 \times (1 - \exp(-(H_{i,0} + H_{i,1-11} + H_{i,12-59}))) \tag{21}$$

2.4   Stage 4: Contextual heterogeneity analysis

While baseline mortality provides a parsimonious threshold criterion, treatment effects may be modified by contextual factors beyond overall mortality levels. Vaccination coverage, malaria burden, and healthcare access could plausibly influence azithromycin's mechanisms of action [19]. We systematically assessed whether such factors (1) explain baseline mortality variation and (2) modify treatment effects beyond mortality-based predictors.

2.4.1   Covariate data sources

We compiled contextual indicators from established global health databases:

- **Vaccination coverage:** DPT1, DPT3, BCG, measles, and polio coverage from IHME gridded surfaces [19]

- **Malaria burden:** Incidence, mortality, and bed net coverage from the Malaria Atlas Project

- **Composite measures:** Geometric mean vaccination coverage and first principal component of vaccination indicators

Covariates were spatially linked to trial clusters using longitude-latitude coordinates, with temporal alignment to trial periods where available. For vaccination indicators, we used immunization gaps (1 - coverage) to ensure consistent interpretation where higher values indicate worse conditions.

2.4.2 Nested model framework

For each covariate x, we fit three nested Poisson generalized linear mixed models to assess explanatory and modifying effects. All models included exposure offsets, natural cubic splines for age, and cluster random intercepts:

Base model (mortality only):

$$\log \mu_{iat} = \log E_{iat} + \beta_0 + f(a) + b_i \tag{22}$$

Covariate model (explanatory effect):

$$\log \mu_{iat} = \log E_{iat} + \beta_0 + \beta_1 x_i + f(a) + b_i \tag{23}$$

Interaction model (modifying effect):

$$\log \mu_{iat} = \log E_{iat} + \beta_0 + \beta_1 x_i + \beta_2 T_{it} + \beta_3 (x_i \times T_{it}) + f(a) + b_i \tag{24}$$

where:

μ_{iat} = expected deaths for cluster i, age group a, treatment arm t

E_{iat} = person-time exposure (offset)

x_i = standardized covariate value for cluster i

T_{it} = treatment indicator (1 if azithromycin, 0 if placebo)

$f(a)$ = natural cubic spline with knots at 5.5, 11.5, 23.5, 41.5 months

b_i = cluster random intercept $\sim \text{Normal}(0, \sigma_b^2)$

β_0 = intercept

β_1 = covariate main effect

β_2 = treatment main effect

β_3 = covariate-treatment interaction effect

2.4.3 Statistical inference

We used likelihood ratio tests to compare nested models:

- **Explanatory test:** Covariate vs. Base model tests whether $\beta_1 \neq 0$

- **Modifying test:** Interaction vs. Covariate model tests whether $\beta_3 \neq 0$

Models were fit using maximum likelihood with robust variance estimation to account for clustering and overdispersion.

2.5 Stage 5: Cessation threshold estimation

Our primary objective was identifying mortality levels where azithromycin's survival benefit disappears, providing empirical cessation criteria. This requires modeling how treatment effects vary continuously with baseline mortality rather than simply testing for effect modification [4].

2.5.1 Interaction model specification

We specified treatment-mortality interaction models separately for U5MR and IMR outcomes:

$$\log \mu_i = \alpha_{g[i]} + \beta_m \log M_i + T_i(\beta_t + \beta_{tm} \log M_i) + \log E_i, \tag{25}$$

where:

μ_i = expected deaths in cluster i

$\alpha_{g[i]}$ = trial-country group intercept for cluster i

M_i = baseline mortality (U5MR or IMR) for cluster i

T_i = treatment assignment indicator

β_m = baseline mortality slope (log scale)

β_t = treatment main effect

β_{tm} = treatment-mortality interaction coefficient

E_i = person-time exposure (offset)

The interaction term $\beta_{tm} \log M_i$ allows treatment effects to vary continuously with baseline mortality. Crucially, we incorporated uncertainty in baseline mortality estimates by sampling M_i values from their posterior distributions (Stage 3) rather than treating them as fixed quantities.

2.5.2 Analytical threshold derivation

Treatment effects become null when the treatment coefficient equals zero:

$$\beta_t + \beta_{tm} \log M^* = 0 \tag{26}$$

Solving for the threshold mortality level:

$$M^* = \exp\left(-\frac{\beta_t}{\beta_{tm}}\right) \tag{27}$$

For each MCMC posterior draw j, we computed:

$$M_j^* = \exp\left(-\frac{\beta_{t,j}}{\beta_{tm,j}}\right) \tag{28}$$

This yields full posterior distributions for cessation thresholds, enabling principled uncertainty quantification. However, this approach can become unstable when $\beta_{tm,j} \approx 0$, necessitating validation through alternative methods.

2.5.3 Crossing-based threshold validation

As a robust alternative, we implemented a crossing-based approach that directly identifies mortality levels where predicted treatment and placebo rates intersect. For each posterior draw, we computed predicted death rates across a dense mortality grid:

$$\lambda_{\text{placebo}}(M) = \exp(\alpha + \beta_m \log M) \tag{29}$$

$$\lambda_{\text{treatment}}(M) = \exp(\alpha + \beta_m \log M + \beta_t + \beta_{tm} \log M) \tag{30}$$

We identified crossing points M^* where:

$$\lambda_{\text{treatment}}(M^*) = \lambda_{\text{placebo}}(M^*) \tag{31}$$

using linear interpolation between grid points to ensure precise threshold estimates.

This approach provides direct visualization of treatment benefit disappearance and serves as validation for analytical estimates. Both methods incorporate full posterior uncertainty, yielding credible intervals for cessation thresholds that reflect both parameter estimation uncertainty and baseline mortality estimation uncertainty.

2.6   Sensitivity analysis

Threshold estimation presents unique challenges for robustness assessment. Unlike standard effect size estimation, thresholds depend on precise identification of mortality levels where treatment effects become null, making them potentially sensitive to modeling assumptions, influential observations, and computational approaches. We implemented a comprehensive sensitivity analysis to ensure our cessation recommendations are robust across alternative specifications and data subsets.

2.6.1   Alternative modeling frameworks

We compared threshold estimates across three methodologically distinct approaches, each representing different assumptions about uncertainty propagation and computational implementation:

Stan variable baseline (primary approach):

$$M_i \sim \text{Posterior distribution from Stage 3 spatial model} \tag{32}$$

This approach treats baseline mortality as uncertain, sampling $M_{i,j}$ values from full posterior distributions for each MCMC iteration j. This properly propagates spatial modeling uncertainty into threshold estimation but requires careful handling of nested uncertainty structures.

Stan fixed baseline (alternative):

$$M_i = \mathrm{E}[\text{Posterior}_i] \quad \text{(point estimates)} \tag{33}$$

Uses posterior means from spatial modeling as fixed covariates, eliminating baseline uncertainty propagation. This approach tests whether uncertainty in mortality estimates substantially affects threshold inference while maintaining the interaction model structure.

INLA implementation (computational validation):

$$M_i = \hat{M}_i \quad \text{(empirical estimates)} \tag{34}$$

Independent implementation using Integrated Nested Laplace Approximation [20] with empirically derived mortality estimates. This provides computational cross-validation using a fundamentally different inference engine while maintaining equivalent model structures.

2.6.2 Sequential exclusion analysis

To assess sensitivity to influential trial sites, we systematically excluded each major trial-country combination and re-estimated thresholds using the INLA framework:

- **Niger MORDOR I/II exclusion:** Tests sensitivity to the largest, highest-mortality trial component (594 clusters, 5,253 deaths)

- **Niger AVENIR exclusion:** Assesses influence of the most extensive geographic coverage (2,158 clusters across Niger)

- **Tanzania MORDOR I exclusion:** Evaluates impact of the lowest-mortality trial setting (360 deaths, 613 clusters)

- **Malawi MORDOR I exclusion:** Tests influence of intermediate mortality levels (304 clusters, 1,044 deaths)

- **Burkina Faso CHAT exclusion:** Assesses impact of the most recent trial data (285 clusters, 1,086 deaths)

This systematic exclusion approach tests whether threshold estimates are driven by specific high-leverage trial locations or represent consistent patterns across diverse epidemiological settings. Each exclusion removes approximately 7-55% of total clusters and 3-45% of observed deaths, providing substantial perturbations to test robustness.

2.6.3 Threshold calculation method comparison

For each modeling scenario described above, we computed cessation thresholds using both analytical and crossing-based approaches (detailed in Stage 5):

Analytical method:

$$M_j^* = \exp\left(-\frac{\beta_{t,j}}{\beta_{tm,j}}\right) \tag{35}$$

Crossing-based method:

$$\lambda_{\text{treatment}}(M_j^*) = \lambda_{\text{placebo}}(M_j^*) \tag{36}$$

This dual-method approach serves multiple purposes: (1) cross-validation of computational implementation, (2) assessment of sensitivity to numerical instabilities when $\beta_{tm,j} \approx 0$, and (3) evaluation of threshold estimate precision across different identification strategies.

2.6.4 Sensitivity metrics and evaluation criteria

We assessed robustness using multiple complementary measures:

- **Central tendency stability:** Comparison of posterior means across scenarios

- **Uncertainty propagation:** Assessment of credible interval overlap between approaches

- **Tail behavior:** Evaluation of extreme percentiles (2.5th, 97.5th) across methods

- **Convergence diagnostics:** $\hat{R}$ statistics and effective sample sizes for all implementations

Results showing consistency across these metrics provide confidence that threshold recommendations are robust to reasonable alternative specifications and not driven by particular modeling choices or influential observations.

3 Results

3.1 Study population

Our dataset comprised 3,954 clusters across five trial-country combinations, covering 1.63 million children who contributed 794,318 person-years of follow-up. During this time, 11,580 deaths were recorded (Table 1). These figures closely match denominators reported in original trial publications [2, 21, 22], confirming data completeness.

Table 1: *Sample characteristics by trial–country setting, ages 1–59 months*

Trial phase	Country	Clusters (n)	Records (n)	Deaths (n)	Person-years
AVENIR	Niger	2,158	619,228	3,837	298,683
CHAT	Burkina Faso	285	237,434	1,086	119,139
MORDOR I	Malawi	304	240,384	1,044	108,009
MORDOR I	Tanzania	613	131,095	360	63,127
MORDOR I/II	Niger	594	400,111	5,253	205,360
Total		3,954	1,628,252	11,580	794,318

3.2 Baseline mortality patterns

Our spatial modeling successfully estimated mortality for all clusters with appropriate uncertainty. Figure 1 shows how this worked: orange distributions are raw cluster estimates (quite variable), while blue distributions show our final estimates after spatial smoothing (more stable). We see how the posterior shrinks towards the prior (given by the dashed line and shaded area and also in the subtitle)—reducing extreme estimates while maintaining genuine variation between locations. This gives us reliable mortality estimates that reflect both local conditions and broader patterns.
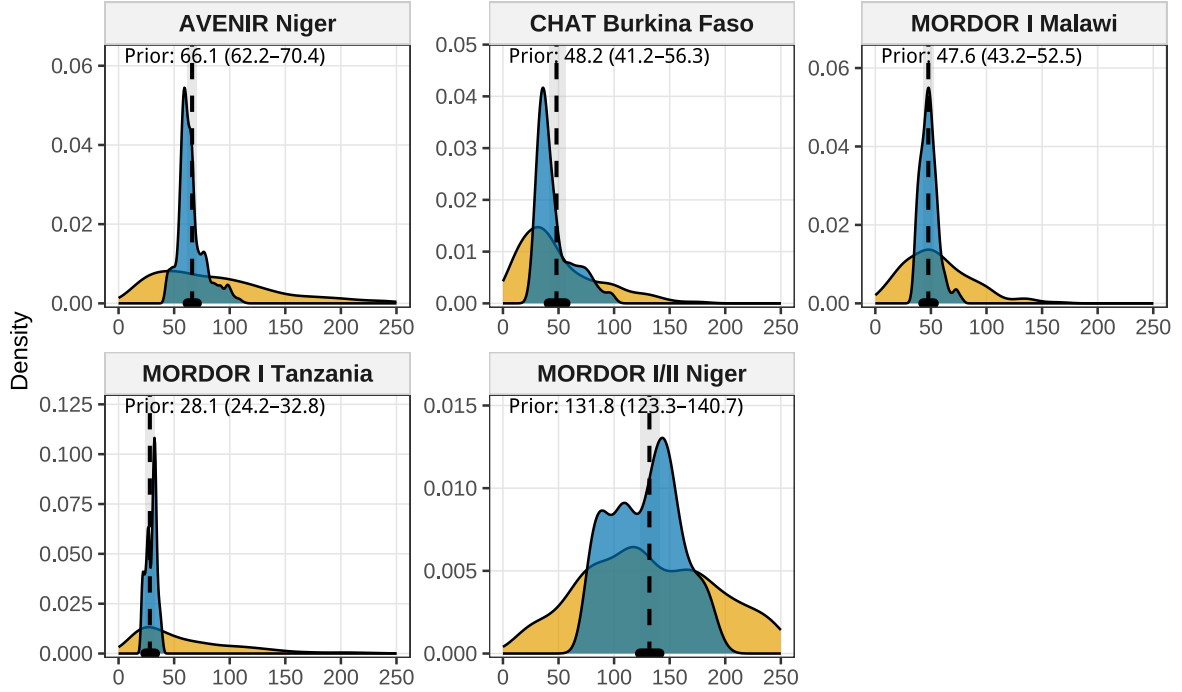
Figure 1: *Spatial modeling reduced wild variability while preserving real differences*

SPDE validation. To validate our HSGP approach, we implemented an alternative specification using stochastic partial differential equations (SPDE) via INLA [23, 20]. The SPDE approach represents the spatial field $u(\mathbf{s})$ as the solution to

$$(\kappa^2 - \Delta)^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s}),$$

where Δ is the Laplacian operator, $\mathcal{W}(\mathbf{s})$ is spatial white noise, and $\alpha = 2$ yields a Matérn covariance function with smoothness $\nu = 1$. We constructed triangulated meshes over cluster locations and fit negative binomial models to placebo-only observations:

$$y_i \sim \text{NegBin}(\mu_i, \phi), \qquad \log \mu_i = \log E_i + \beta_0 + u(\mathbf{s}_i),$$

with penalized complexity priors on range and marginal variance parameters [24].

The SPDE implementation yielded cluster-level mortality estimates highly correlated with our HSGP + HS results across all trial-country combinations (Figure 2). The SPDE approach pro-

vides a useful benchmark for validating our HSGP with horseshoe results. As shown in Figure 2, posterior distributions from the SPDE (green) and HSGP (blue) align closely across trial–countries, both in central tendency and distributional shape. Because SPDE relies on a mesh-based approximation of Matérn fields, independent of the basis-function construction in HSGP, the agreement confirms that the HSGP estimates are not artefacts of model specification but instead capture the underlying mortality patterns robustly.
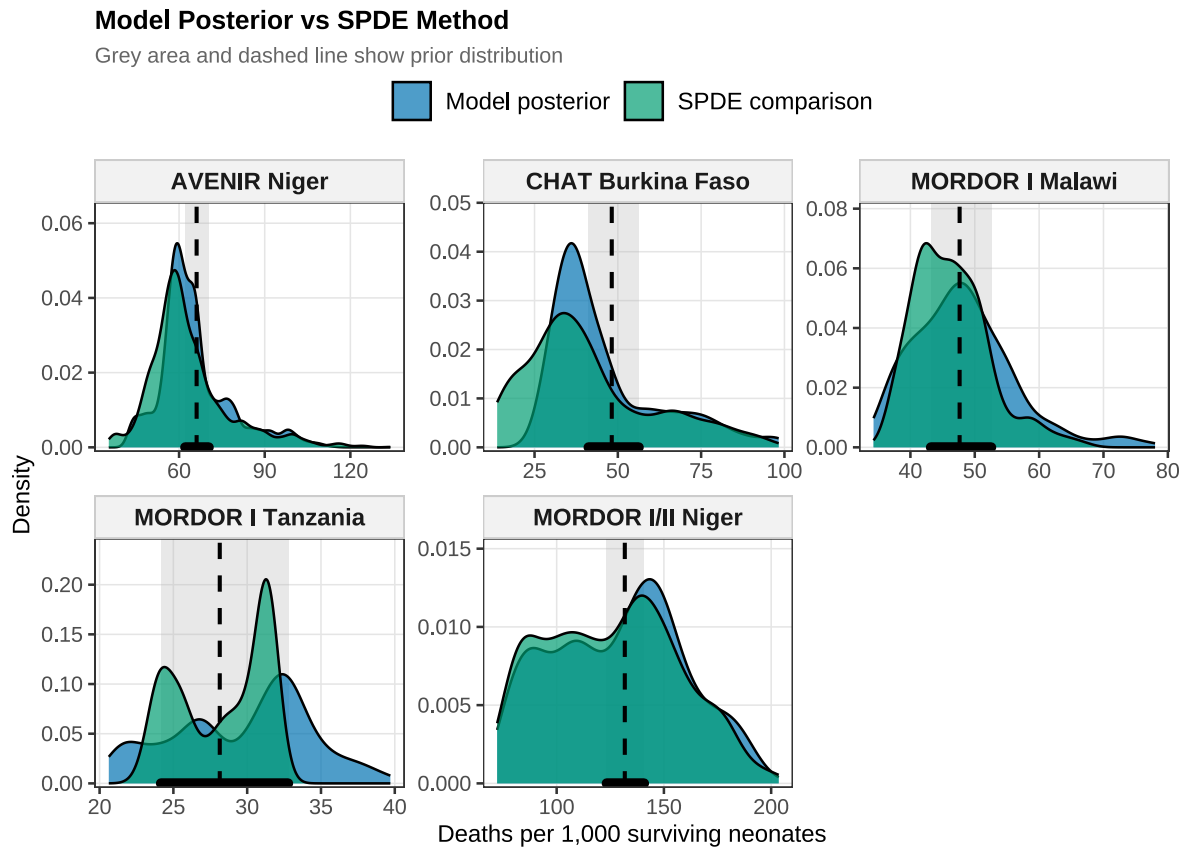
Figure 2: *HSGP + HS produces estimates comparable to SPDE*

However, the SPDE approach requires multi-stage modeling to propagate uncertainty through age decomposition and neonatal bridging, whereas our HSGP formulation generates predictions directly in deaths space within a single Stan model. This unified approach preserves full posterior uncertainty through all transformations from cluster-specific hazards to demographic indicators, making it preferable for threshold estimation where uncertainty quantification is critical.

3.2.1 Mapping to demographic indicators

Translation to standard demographic indicators also worked well. Table 2 shows mortality estimates by trial-country and treatment group. IMR:U5MR ratios stay within expected bounds (0.5–0.65), validating our approach. Treatment and placebo groups within each trial-country show similar estimates, supporting successful prediction out-of-sample (only placebo where used to estimate baseline).

Table 2: *Baseline mortality estimates by trial-country and treatment arm*

Trial–Country	Treatment	IMR (mean and SD)	U5MR (mean and SD)	Ratio
AVENIR Niger	Placebo	45.7 (6.40)	88.5 (14.73)	0.517
AVENIR Niger	Azithromycin	45.6 (5.76)	88.2 (13.23)	0.517
CHAT Burkina Faso	Placebo	32.8 (7.35)	67.7 (19.91)	0.485
CHAT Burkina Faso	Azithromycin	32.4 (6.77)	66.5 (18.21)	0.487
MORDOR I Malawi	Placebo	41.3 (5.01)	71.1 (10.27)	0.581
MORDOR I Malawi	Azithromycin	41.6 (4.16)	71.6 (8.53)	0.581
MORDOR I Tanzania	Placebo	30.6 (3.20)	47.7 (5.88)	0.640
MORDOR I Tanzania	Azithromycin	30.8 (3.12)	48.2 (5.74)	0.639
MORDOR I/II Niger	Placebo	76.4 (13.77)	160.0 (32.34)	0.478
MORDOR I/II Niger	Azithromycin	74.1 (14.08)	154.5 (33.05)	0.479

3.3 Contextual heterogeneity

We systematically tested whether factors beyond baseline mortality might predict treatment effects. As expected, vaccination coverage and malaria burden were correlated with mortality (Figure 3). Places with poor vaccination coverage showed strong correlations with higher mortality, while malaria indicators showed more moderate associations.

The key question was whether these factors modified treatment effects beyond their correlation with baseline mortality. Figure 4 shows the results clearly. The left panel confirms that vaccination and malaria factors predict baseline mortality (many significant results). The right panel shows treatment interactions—and there's little systematic pattern.

Out of 45 treatment interaction tests, only 5 were statistically significant—about what you'd expect by chance. The patterns weren't consistent across sites either. This tells us that while factors like vaccination and malaria affect baseline risk, they don't systematically change how well azithromycin works beyond what baseline mortality already captures.
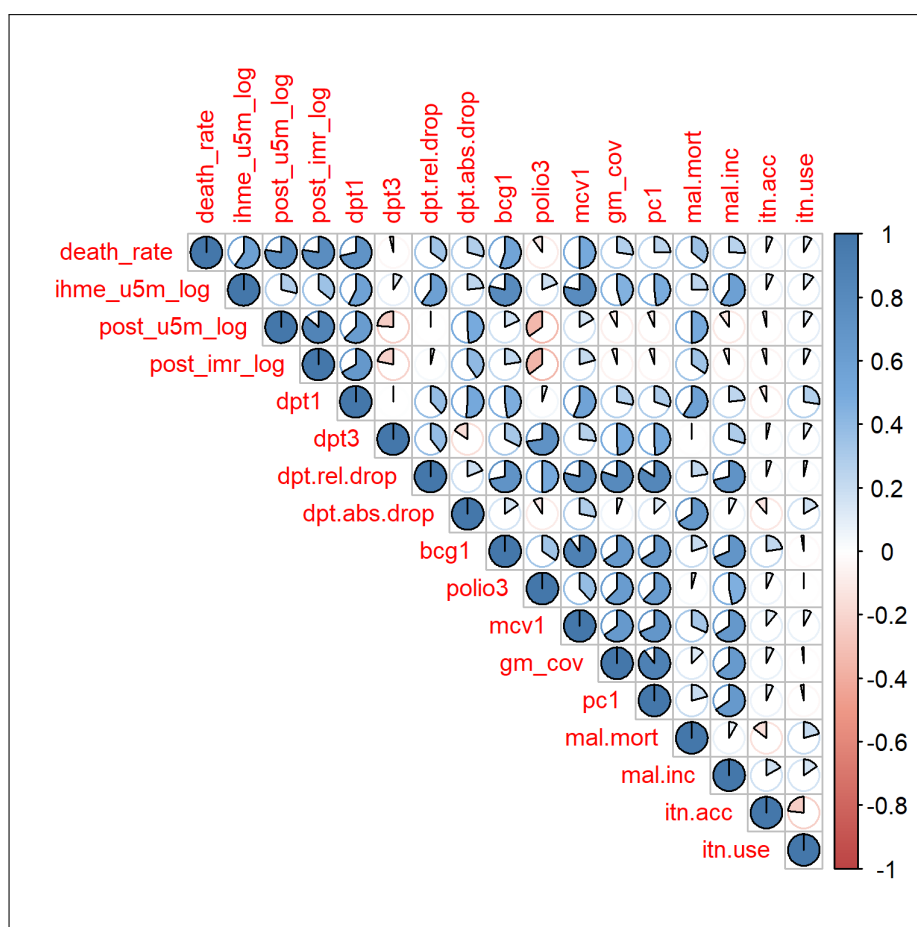
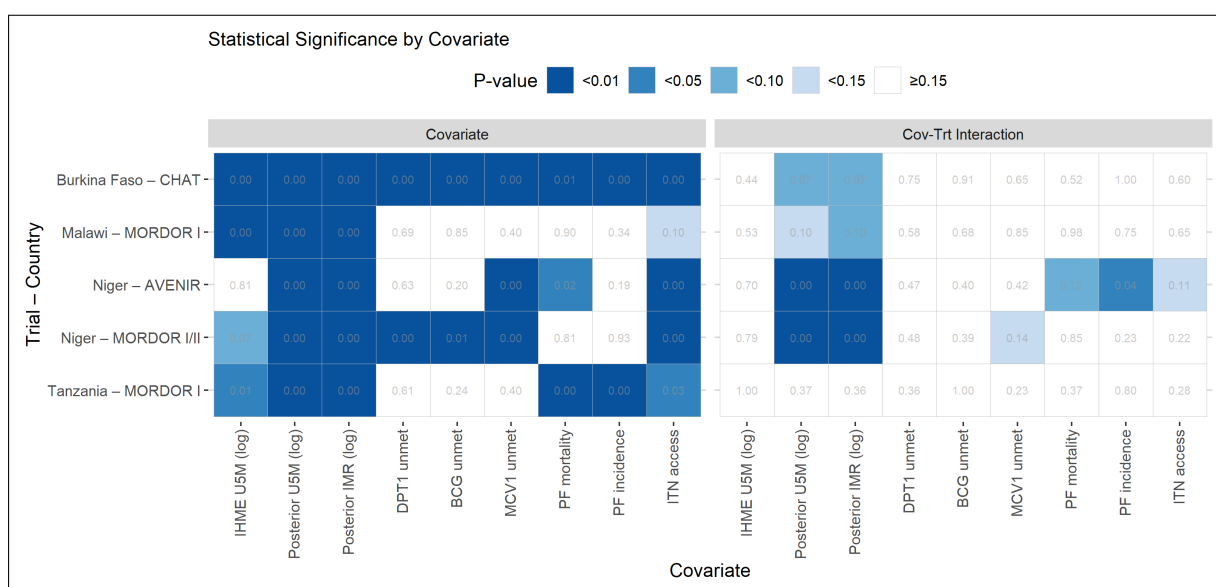Figure 3: *Contextual factors correlate with baseline mortality as expected*

Figure 4: *Contextual factors predict mortality but don't systematically modify treatment effects*

These findings support focusing on baseline mortality as the primary criterion for cessation decisions.

3.4 The cessation thresholds

Our treatment-mortality interaction models successfully converged for both IMR and U5MR, yielding the first empirically grounded cessation thresholds. Table 3 shows our main results.

Table 3: *Data-driven cessation thresholds for azithromycin MDA programs*

Model	Threshold (per 1,000)
Under-five mortality (U5MR)	49.4 (35.8-72.0)
Infant mortality (IMR)	27.3 (20.3-37.9)

The U5MR-based threshold is 49.4 (35.8-72.0) per 1,000 live births; the IMR-based threshold is 27.3 (20.3-37.9) per 1,000 live births. Figure 5 shows both the uncertainty around these thresholds (left panel) and the statistical evidence supporting them (right panel).
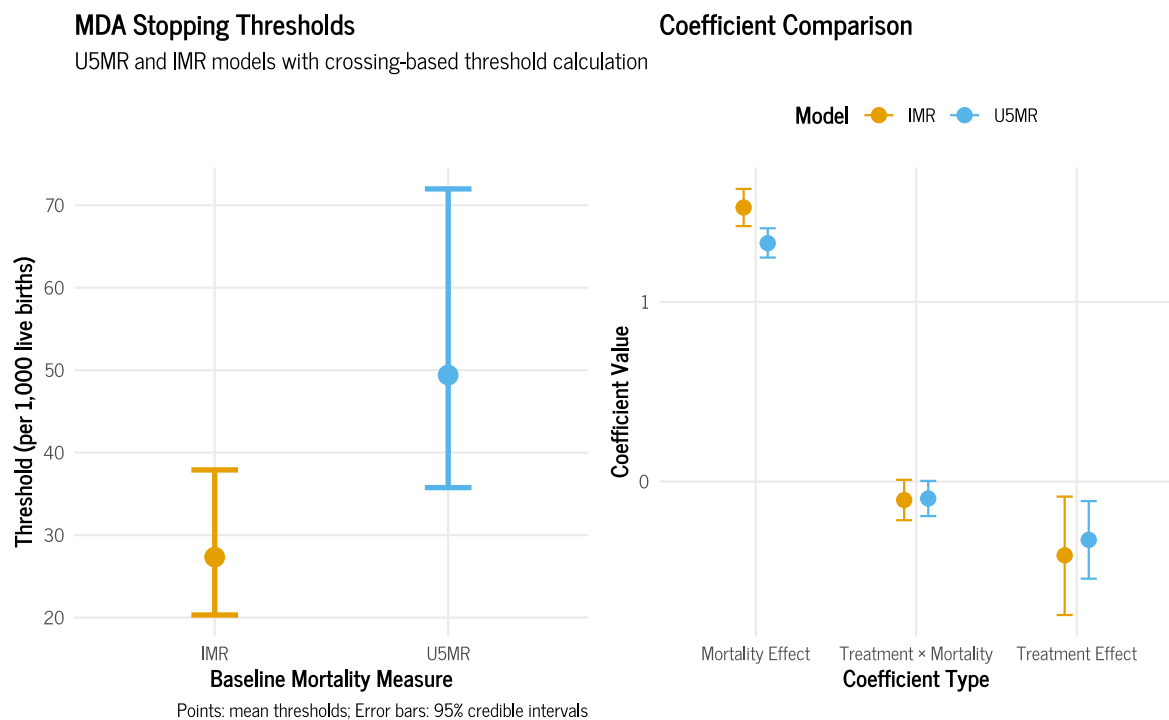
Figure 5: *Cessation thresholds with uncertainty and supporting evidence*

The model coefficients show clear evidence for treatment-mortality interactions, with negative interaction terms indicating diminishing benefits as baseline mortality decreases—exactly what we need for meaningful threshold identification.

The prediction curves (Figure 6 show how this works in practice. At high mortality levels, treatment and placebo groups have very different predicted death rates—azithromycin provides substantial benefit. As baseline mortality drops, these curves converge until they cross at our estimated thresholds. Below these crossing points, azithromycin provides no meaningful benefit.
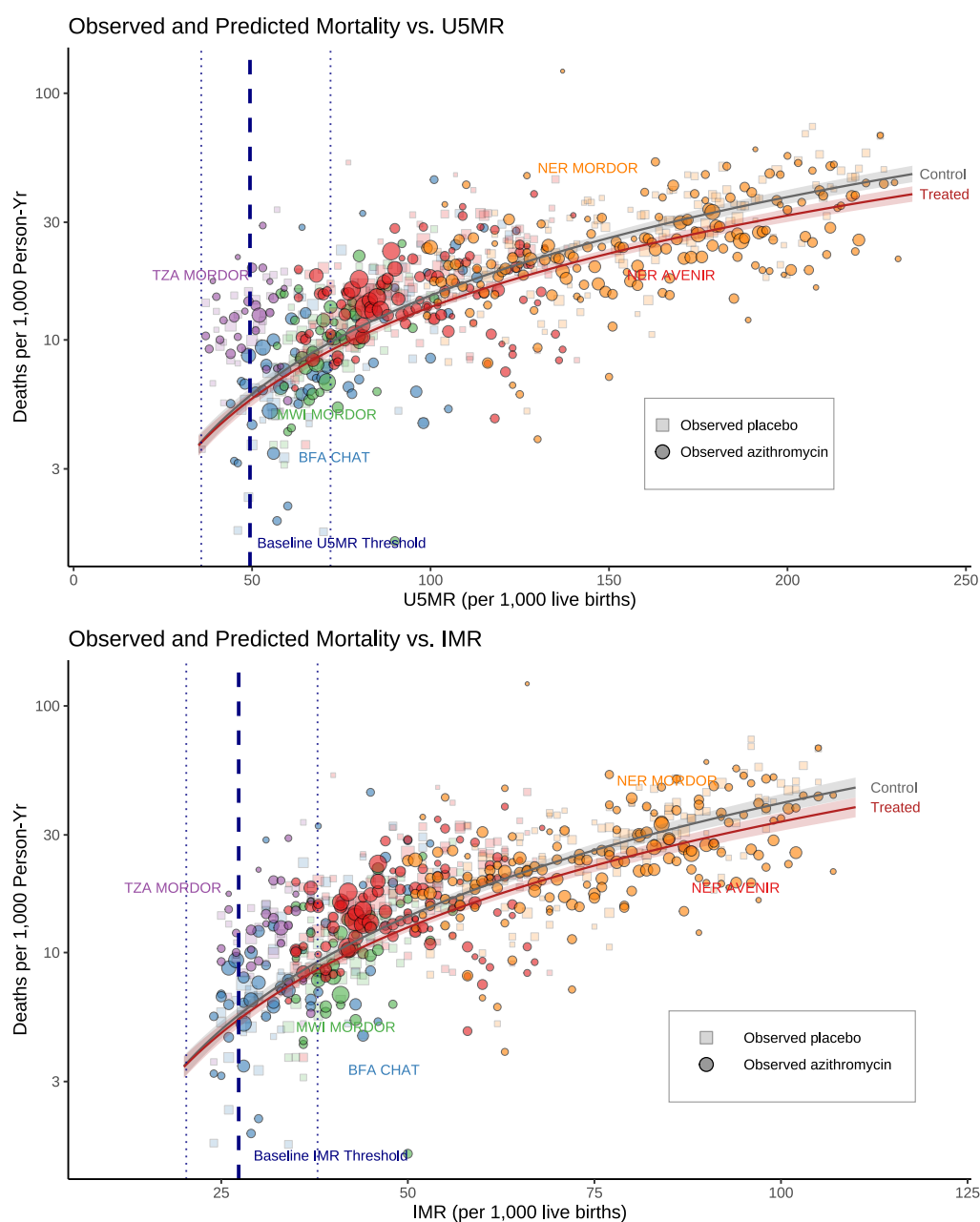

Figure 6: *Treatment benefits vanish as baseline U5MR/IMR increase*

The uncertainty ranges around these thresholds acknowledge the limits of available data while providing actionable guidance for program managers.

3.5 Sensitivity analysis

To assess robustness, we tested our threshold estimates across multiple modeling approaches and data subsets. Figure 7 shows threshold distributions using the crossing method across eight scenarios, while Figure 8 shows results from the analytical approach.
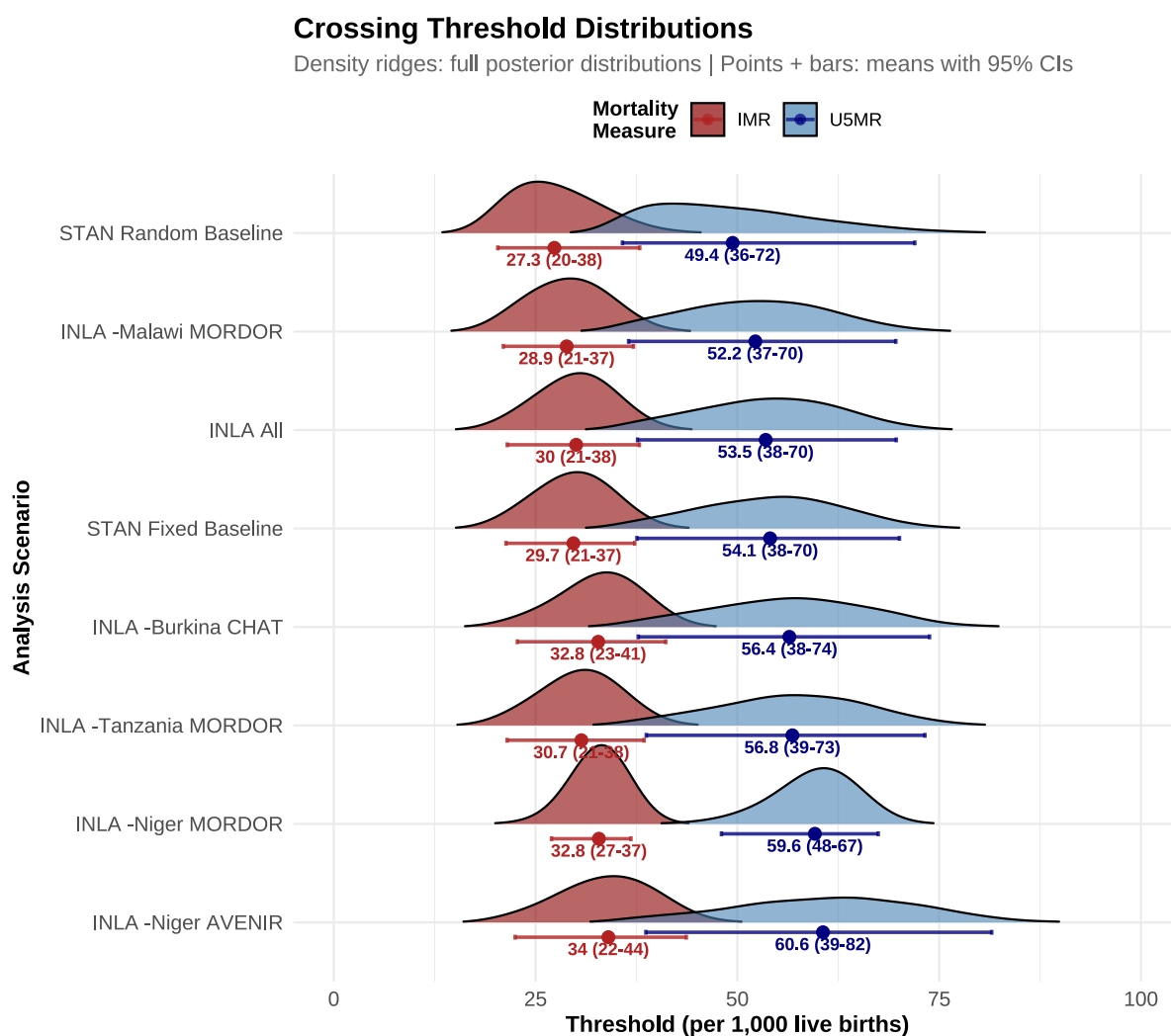

Figure 7: *Crossing-based threshold distributions across modeling scenarios*

Table 4: *Threshold estimates across modeling scenarios and calculation methods*

| scenario	U5MR_analytical	U5MR_crossing	IMR_analytical	IMR_crossing
STAN Random Baseline	35.9 (7.2-68.3)	49.4 (35.8-72.0)	21.0 (6.2-36.6)	27.3 (20.3-37.9)
INLA -Malawi MORDOR	49.5 (22.8-69.1)	52.2 (36.5-69.6)	27.5 (14.2-36.9)	28.9 (21.0-37.1)
INLA All	52.1 (29.1-69.5)	53.5 (37.6-69.7)	29.4 (17.3-37.7)	30.0 (21.5-37.9)
STAN Fixed Baseline	52.8 (30.9-69.9)	54.1 (37.6-70.1)	29.0 (17.4-37.2)	29.7 (21.3-37.3)
INLA -Burkina CHAT	54.8 (29.3-73.4)	56.4 (37.7-73.8)	32.2 (19.3-41.1)	32.8 (22.7-41.1)
INLA -Tanzania MORDOR	55.9 (33.6-73.1)	56.8 (38.7-73.2)	30.0 (17.9-38.4)	30.7 (21.5-38.4)
INLA -Niger MORDOR	59.6 (47.9-67.4)	59.6 (48.0-67.4)	32.8 (26.9-36.8)	32.8 (27.0-36.8)
INLA -Niger AVENIR	59.1 (30.2-81.3)	60.6 (38.7-81.5)	33.4 (18.4-43.6)	34.0 (22.5-43.7)
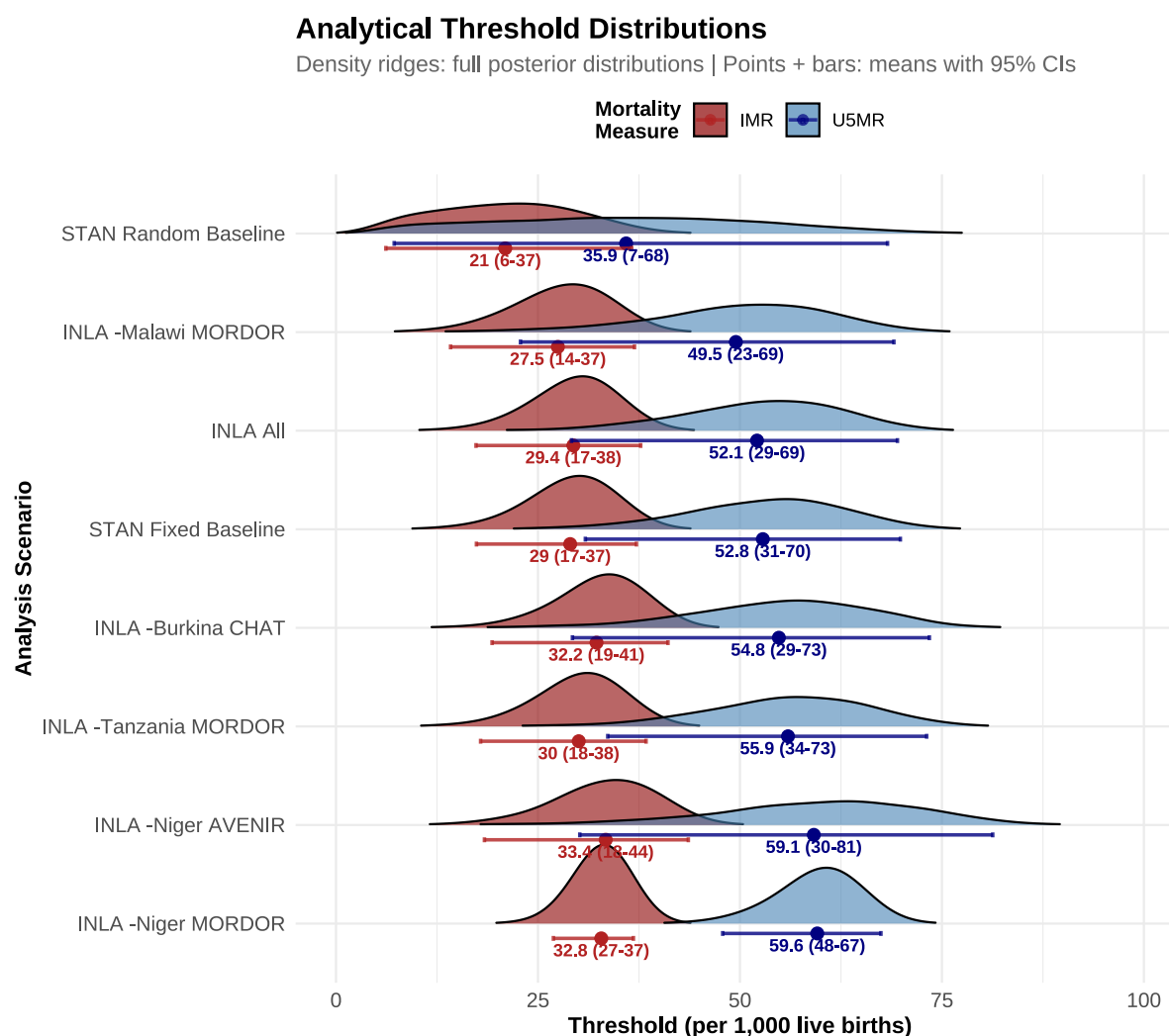
Figure 8: *Analytical threshold distributions across modeling scenarios*

Table 4 summarizes threshold estimates across all scenarios and methods. The crossing method shows consistent estimates across different modeling approaches, while the analytical method exhibits some variability in certain scenarios. Overall, estimates demonstrate reasonable consistency across alternative specifications and data subsets.

4 Discussion

This study provides the first data-driven cessation thresholds for azithromycin MDA programs. We found that benefits disappear when U5MR falls below approximately 49.4 (35.8-72.0) per 1,000 or IMR falls below approximately 27.3 (20.3-37.9) per 1,000 live births.

Our approach successfully translated trial hazards into demographic indicators that policymakers understand. The four-stage Bayesian framework allowed us to combine sparse cluster-level observations with spatial structure and demographic constraints, yielding reliable mortality estimates across diverse settings.

The finding that contextual factors beyond baseline mortality don't systematically modify treatment effects simplifies program guidance. While vaccination coverage and malaria burden correlate with mortality, they don't change how well azithromycin works beyond what baseline mortality already tells us. This supports a parsimonious approach focused on mortality thresholds rather than complex multi-factor criteria.

These thresholds compare favorably to current WHO recommendations. The WHO suggested cessation when IMR falls below 60 per 1,000 or U5MR falls below 80 per 1,000 [6]. Our empirically derived thresholds are lower, suggesting programs might continue longer than current guidance would suggest—but with the important caveat that resistance monitoring remains essential.

4.1 Limitations

Several limitations should be acknowledged. We modeled baseline mortality as uncertain rather than fixed, properly accounting for estimation error, but results depend on the quality of spatial modeling assumptions. Alternative specifications yield similar crossings with wider intervals, supporting robustness. The inference on context is based on a limited set of variables tested and the precision of these in time and space may not be high enough to detect an association (rather than that being a real result).

We don't model antimicrobial resistance dynamics directly. Programs should interpret thresholds alongside local surveillance data and resistance monitoring. Our findings reflect the specific trial settings analyzed. While we included diverse epidemiological contexts, generalization to other settings requires careful consideration of local factors.

References

[1] United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). Levels & Trends in Child Mortality: Report 2024. Technical report, United Nations Children's Fund (UNICEF), New York, 2024. URL https://data.unicef.org/wp-content/uploads/2025/05/UNIGME-2024-Child-Mortality-Report_13-May.pdf.

[2] Jeremy D. Keenan, Robin L. Bailey, Sheila K. West, et al. Azithromycin to reduce childhood mortality in sub-saharan africa. *New England Journal of Medicine*, 378(17):1583–1592, 2018. doi:10.1056/NEJMoa1715474.

[3] Jeremy D. Keenan, Ahmed M. Arzika, Ramatou Maliki, et al. Longer-term assessment of azithromycin for reducing childhood mortality in africa. *New England Journal of Medicine*, 380(23):2207–2214, 2019. doi:10.1056/NEJMoa1817213.

[4] Assaf P. Oron, Roy Burstein, Laina D. Mercer, Ahmed M. Arzika, Khumbo Kalua, Zakayo Mrango, Sheila K. West, Robin L. Bailey, Travis C. Porco, and Thomas M. Lietman. Effect modification by baseline mortality in the MORDOR azithromycin trial. *American Journal of Tropical Medicine and Hygiene*, 103(3):1295–1300, 2020. doi:10.4269/ajtmh.18-1004.

[5] Akuzike Kalizang'oma, Jia Mun Chan, Khumbo Kalua, Farouck Bonomali, Comfort Brown, Jacqueline Msefula, David Chaima, Lyson Samikwa, Harry Meleke, John D Hart, et al. Long-term effects of azithromycin mass administration to reduce childhood mortality on streptococcus pneumoniae antimicrobial resistance: a population-based, cross-sectional, follow-up carriage survey. *The Lancet Infectious Diseases*, 2025.

[6] World Health Organization. Who guideline on the use of azithromycin mass drug administration for trachoma elimination and for reduction of childhood mortality in settings with high child mortality. Technical report, WHO, Geneva, 2020.

[7] Violeta Jimenez, Huub C Gelderblom, Rebecca Mann Flueckiger, Paul M Emerson, and Danny Haddad. Mass drug administration for trachoma: how long is not long enough? *PLoS neglected tropical diseases*, 9(3):e0003610, 2015.

[8] Hamidah Mahmud, Emma Landskroner, Abdou Amza, Solomon Aragie, William W Godwin, Anna de Hostos Barth, Kieran S O'Brien, Thomas M Lietman, and Catherine E Oldenburg. Stopping azithromycin mass drug administration for trachoma: A systematic review. *PLoS Neglected Tropical Diseases*, 15(7):e0009491, 2021.

[9] Swarnali Sharma, Morgan E Smith, Shakir Bilal, and Edwin Michael. Evaluating elimination thresholds and stopping criteria for interventions against the vector-borne macroparasitic disease, lymphatic filariasis, using mathematical modelling. *Communications Biology*, 6(1):225, 2023.

[10] Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991. doi:10.1007/BF00116466.

[11] Laina D. Mercer, Jon Wakefield, Athanasios Pantazis, Aniset Lutambi, Honorati Masanja, and Samuel Clark. Space-time smoothing of complex survey data: small area estimation for child mortality. *Annals of Applied Statistics*, 9(4):1889–1905, 2015. doi:10.1214/15-AOAS872.

[12] Jonathan Wakefield, Taylor Okonek, and Jon Pedersen. Small area estimation for disease prevalence mapping. *International Statistical Review*, 88(2):398–418, 2020. doi:10.1111/insr.12400.

[13] Li Liu, Shefali Oza, Daniel Hogan, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. *The Lancet*, 388(10063):3027–3035, 2016. doi:10.1016/S0140-6736(16)31593-8.

[14] Andrea Riebler, Sigrunn H. Sørbye, Daniel Simpson, and Håvard Rue. An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165, 2016. doi:10.1177/0962280216660421.

[15] Arno Solin and Simo Särkkä. Hilbert space methods for reduced-rank gaussian process regression. *Statistics and Computing*, 30(2):419–446, 2020.

[16] Gabriel Riutort-Mayol, Paul-Christian Bürkner, Michael R Andersen, Arno Solin, and Aki Vehtari. Practical hilbert space approximate bayesian gaussian processes for probabilistic programming. *Statistics and Computing*, 33(1):17, 2023.

[17] Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010. doi:10.1093/biomet/asq017.

[18] Juho Piironen and Aki Vehtari. On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In *Proceedings of the 20th International Conference on Artificial*

Intelligence and Statistics, volume 54 of *Proceedings of Machine Learning Research*, pages 905–913. PMLR, 2017. URL http://proceedings.mlr.press/v54/piironen17a.html.

[19] Nick Golding, Roy Burstein, Joshua Longbottom, et al. Mapping under-5 and neonatal mortality in africa, 2000–15: a baseline analysis for the sustainable development goals. *The Lancet*, 390(10108):2171–2182, 2017. doi:10.1016/S0140-6736(17)31758-0.

[20] Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63:1–25, 2015.

[21] Catherine E. Oldenburg, Mamadou Ouattara, Mamadou Bountogo, et al. Mass azithromycin distribution to prevent child mortality in Burkina Faso: The CHAT randomized clinical trial. *JAMA*, 331(6):482–490, 2024. doi:10.1001/jama.2023.27393.

[22] Kieran S O'Brien, Ahmed M Arzika, Abdou Amza, Ramatou Maliki, Bawa Aichatou, Ismael Mamane Bello, Diallo Beidi, Nasser Galo, Naser Harouna, Alio M Karamba, et al. Azithromycin to reduce mortality—an adaptive cluster-randomized trial. *New England Journal of Medicine*, 391(8):699–709, 2024.

[23] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(4): 423–498, 2011.

[24] Geir-Arne Fuglstad, Daniel Simpson, Finn Lindgren, and Håvard Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019. doi:10.1080/01621459.2017.1415907.