**Additional file 2:**

*Bayesian model formulation*

Let Y*jt* be the observed number of deaths (all-cause or malaria-specific) in village *j* at time *t, j=1,2,…n* and *t=1,2,…9* for the model fitted on the annual data. We assume that Y*jt* arises from a negative binomial distribution$, Y\_{jt}\~NB(μ\_{jt},r)$where $μ\_{jt}$, is the expected number of deaths and *r*is the dispersion parameter. We modeled malaria prevalence and other covariates (***X***) by on the log scale of $μ\_{jt}$

$log\left(μ\_{jt}\right)=log\left(N\_{jt}\right)+β\_{0}+\sum\_{1}^{k}β\_{k}X\_{jtk}+ϕ\_{j}$ , *k=1,2,…K*

where $N\_{jt} $the total person time contributed by persons in village *j* at time *t* (in person years, py), $β=(β\_{1},β\_{2},…,β\_{K})^{T}$ are the regression coefficients (malaria prevalence and other predictors) and $ϕ\_{j}$ the village specific spatial effects. We assumed that $ϕ=(ϕ\_{1},ϕ\_{2},…ϕ\_{n})^{T}$ are modeled by a Gaussian process, that is $ϕ \~ MVN(0,σ\_{1}^{2}R)$ and that $R$ is an exponential correlation matrix of the distance between villages, i.e. $R\_{ij}=exp⁡(-ρd\_{kl})$ where $d\_{kl} $is the distance between villages  and , $ρ$ is the rate of *correlation* decay with distance. The minimum distance at which the spatial correlation is less than 5% is called effective range and is defined by the value of $3/ρ$ (1). The $σ\_{1}^{2}$ is the variance of the spatial process. We specified non-informative normal prior distributions with mean zero and large variance for the regression coefficients $β\_{i}\~N(0, 10^{3})$, an inverse gamma prior for $σ\_{1}^{2}$, that is $σ\_{1}^{2} \~IG(2.01, 1.01)$ and a Uniform prior distribution for $ρ$, that is $ρ\~U(a, b)$, where $a$ and $b$ are chosen such as the effective range is within the maximum and minimum distances of the village’s locations.

The Bayesian models were then fitted using Markov Chain Monte Carlo (MCMC) simulation algorithm in OpenBugs version 3.1.2 (Imperial College and Medical Council, London, UK) to estimate model parameters (2). We ran a two chain sampler for 100,000 iterations, discarding the first 10,000 iterations. Convergence was assessed by the Gelman-Rubin diagnostic (3).

*References*

1. Ecker MD, Gelfand AE. Bayesian Variogram Modeling for an Isotropic Spatial Process. J Agric Biol Environ Stat. 1997;2(4):347–69.

2. Gelfand AE, Smith AFM. Sampling-Based Approaches to Calculating Marginal Densities. J Am Stat Assoc. 1990 Jun 1;85(410):398–409.

3. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci. 1992 Nov;7(4):457–72.