SUPPLEMENTARY MATERIAL

Associations and contribution of childhood diseases to fever risk among children less than five years in Uganda

Betty Bukenya Nambuusi[1,2,3], Julius Ssempiira[1,2,3], Fredrick E Makumbi[3], Simon Kasasa[3], Penelope Vounatsou[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Basel, Switzerland

*3*Makerere University School of Public Health, New Mulago Hospital Complex, Kampala, Uganda

Appendix S1. Bayesian geostatistical modeling

Bayesian geostatistical variable selection

A Bayesian geostatistical logistic regression model, adopting a stochastic search variable selection approach, was used to determine the most important predictors of fever prevalence and their functional form [1]. ITN coverage measures were grouped into use and ownership. Only one (or none) ITN measure among those defining ownership and one (or none) ITN measure among those defining use [2] was selected. For each ITN coverage measure X_p in the ownership group, a categorical indicator I_p, was introduced to represent exclusion of the variable from the model ($I_p = 1$), or inclusion of one of the ITN ownership measure i.e. %hh1itn ($I_p = 2$), %hh1itn4two ($I_p = 3$) and %pp1itn $\left(I_p = 4\right)$. A similar definition was adopted for the ITN use coverage measure i.e. exclusion of the variable from the model ($I_p = 1$), inclusion of %ppslept ($I_p = 2$), %chslept ($I_p = 3$) and %itnused $\left(I_p = 4\right)$. The ITN measure with the highest probability of inclusion in each category was included in the final model.

For the environmental/climatic variables except of land cover types, variable selection compared their linear and categorical forms and selected the one that had the highest probability of inclusion or neither of the two forms. The categorical forms were generated based on the quartiles of variables. We introduced an indicator I_p for each environmental/climatic covariate X_p which defines exclusion of the variable from the model ($I_p = 1$), inclusion in a categorical ($I_p = 2$) or linear ($I_p = 3$) form.

For childhood diseases, vaccinations, treatments, health care seeking characteristics, socio-demographics and land cover types, a binary indicator parameter I_p suggesting presence ($I_p = 1$) or absence ($I_p = 0$) of the predictor from the model was introduced. I_p has a probability mass function $\prod_{j=1}^{p} \pi_j^{\delta_j(I_p)}$, where π_j denotes the inclusion probabilities, $j =$

(1,2,3,4), e.g. for ITN coverage measures so that $\sum_{j=1}^{p}\pi_j = 1$ and $\delta_j(.)$ is the Dirac function,

$$\delta_j(I_p) = f(x) = \begin{cases} 1, if\ I_p = j \\ 0, if\ I_p \neq j \end{cases}.$$

For inclusion probabilities of ITN use and ownership, a non-informative Dirichlet distribution was adopted with hyper parameters $\alpha = (1,1,1,1)^T$, that is, $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_{3,}\pi_4)^T \sim Dirichlet(4, \alpha)$. A similar distribution was adopted for the inclusion probabilities of environmental/climatic factors. For childhood diseases, vaccinations, treatments, health care seeking characteristics, socio-demographics and land cover types, a Bernoulli prior with an equal inclusion or exclusion probability was assumed for the indicator i.e. $I_p \sim bern(0.5)$. Also, inverse Gamma priors with parameters (2.01, 1.01) were assumed for the precision hyper parameters τ_p^2. The predictors identified as important are those with posterior inclusion probability greater than or equal to 50% [3].

We assumed a spike and slab prior for regression coefficient β_p corresponding to the corresponding covariate, X_p i.e. for the coefficient β_p of the predictor in linear form we take $\beta_p \sim \big(1 - \delta_1(I_p)\big)N\big(0, u_0\tau_p^2\big) + \delta_1(I_p)N(0, \tau_p^2)$ proposing a non-informative prior for β_p in case X_p is included in the model in a linear form (slab) and an informative normal prior with variance close to zero (i.e. $u_0 = 10^{-3}$) shrinking β_p to zero (spike) if X_p is excluded from the model. Similarly, for the coefficient $\{\beta_{p,l}\}_{l=1,..,L}$ corresponding to the categorical form of $\underline{X_p}$ with L categories, we assume that $\beta_{p,l} \sim \delta_2(I_p)N\big(0, \tau_{p,l}^2\big) + (1 - \delta_2)N\big(0, \vartheta_0\tau_{p,l}^2\big)$

Bayesian geostatistical logistic regression model with spatially varying effects of childhood diseases

A Bayesian geostatistical logistic regression model [4] was fitted to quantify the effects of childhood diseases on the fever prevalence. The model included spatially varying coefficients for childhood diseases adjusted for socio-demographic factors, vaccinations, health care seeking characteristics, treatments, ITN use and ownership and climatic/environmental

factors. The model assesses the effects of childhood diseases at a regional level using spatially varying coefficients [2] and is formulated assuming a conditional autoregressive (CAR) prior distribution [5]. The CAR introduces a neighbour-based spatial structure for the regression coefficients for each childhood disease effect [6]. Neighbours were defined as the adjacent areas for each region. To adjust for spatial correlation present in the fever prevalence due to similar exposure effect in neighbouring clusters, cluster-specific random effects were introduced into the model. The cluster random effects were assumed to arise from a Gaussian stationary process with a covariance matrix capturing correlation between any pair of cluster locations as a function of their interlocation distances.

Let Y_{ij} be the binary outcome for child i at location s_j taking values 1 and 0 when fever is present or absent respectively, $\boldsymbol{X_j}(s_j)$ be the vector of socio-demographic factors, vaccinations, health care seeking characteristics, treatments, ITN use and ownership and climatic/environmental factors and $Z_d(s_j)$ be the prevalence of disease d at location s_j.

Y_{ij} is assumed to follow a Bernoulli distribution $Y_{ij} \sim Ber(p_{ij})$ and is related to its predictors using a logistic regression model as follows;

$logit(p_{ij}) = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{X_j}(s_j) + \sum_{d=1}^{D} \left(b_d + \varepsilon_{dk(j)}\right)^T Z_d(s_j) + W(s_j) + v_j$ where p_{ij} is the presence or absence of fever of child i at location s_j, $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_p)$ is the vector of regression coefficients with $\exp(\beta_l), l = 1, \ldots p$, corresponding to the odds ratio. $W(s_j)$ is a cluster-specific random frailty which captures spatial correlation in the fever prevalence due to similar exposure effect in neighbouring clusters. We modeled $\boldsymbol{W}(s) = (W(s_1), W(s_2), \ldots, W(s_m))^T$ by a Gaussian process, i.e. $\boldsymbol{W}(s) \sim N(0, \Sigma)$, where Σ is the variance-covariance matrix and each element is defined by an exponential correlation function of the distance d_{kl} between locations s_k and s_l, that is $\Sigma_{kl} = \sigma^2 \exp(-d_{kl}\rho)$ [4]. The parameter σ^2 gives the variance of the spatial process and ρ is a smoothing parameter that controls the rate of correlation decay with distance. For the exponential correlation

function, $\frac{-\log(0.05)}{\rho}$ determines the distance at which the correlation drops to 0.05 (i.e. effective range of spatial process). Non-spatial variation is estimated by the random effects v_j, assumed independent and normally distributed with mean 0 and variance σ_v^2 .

Our model assumed that the relation between childhood diseases and fever prevalence varied across regions by including disease specific spatially varying coefficients, $b_d + \varepsilon_{kd}$, where b_d is the effect of the disease $d = 1, \dots D$ on fever prevalence at country (national) level and $\boldsymbol{\varepsilon_d} = (\varepsilon_{d1}, \dots, \varepsilon_{dk})^{\boldsymbol{T}}$ are the varying effects at regional (sub-national) levels $k = 1, \dots K$ with $k(j)$ indicating the region k corresponding to the location s_j.We introduced spatial dependence among the regions via a conditional autoregressive (CAR) prior for $\boldsymbol{\varepsilon_d}$, that is $\boldsymbol{\varepsilon_d} \sim N(\mathbf{0}, \Omega_d)$ with $\Omega_d = \sigma_d^2(I - \gamma C)^{-1}\Delta$. σ_d^2 is the variance of spatially varying disease effects, Δ is a diagonal matrix with entries $\Delta_{kk} = {g_k}^{-1}$where g_k is the number of neighbours of region k, γ measures overall spatial dependence and C is the adjacency matrix with normalized entries that is $C_{kl} = \omega_{kl}/g_k$, ω_{kl} is 1 if region k neighbors l and 0 otherwise [4].

Model specification was completed by assigning prior distributions to model parameters. We assumed inverse gamma priors for all spatial variances with known parameters, i.e. $\sigma^2, \sigma_d^2 \sim IG(2.01, 1.01)$, a uniform prior distribution for $\rho \sim U(a, b)$, where a and b chosen such that the effective range is within the maximum and minimum distances of the observed locations [7] and a uniform prior for $\gamma \sim U(\lambda_1^{-1}, \lambda_2^{-1})$ where λ_1, λ_2 are the smallest and largest eigenvalue of $\Delta^{-1/2}C\Delta^{1/2}$ [4]. Non-informative normal priors were adopted for the regression coefficients $\beta_l, b_d \sim N(0, 10^3)$ for $l = 1, \dots p$ and $d = 1, \dots D$. The joint posterior distribution of the model is given by

$\prod_{\boldsymbol{j}} [\, Y_i(s_j) \,|\boldsymbol{\beta}, b_d, W(s_j)\, \varepsilon_{dk}, \boldsymbol{X_j}(s_i), \boldsymbol{Z_d}(s_j)][\, \boldsymbol{W}(s)|\sigma^2, \rho][\boldsymbol{\varepsilon_d}|\, \sigma_d^2, \gamma][\boldsymbol{\beta}, b_d, \sigma^2, \sigma_d^2, \gamma, \rho]$. To select priors, we made assumptions based on the direction an estimate could take. For

example, for regression coefficients, a normal distribution was assumed since these estimates could take on negative or positive values. A gamma distribution was assumed for variances as these could only take on positive values.

Model parameters were estimated using Markov Chain Monte Carlo simulation [1]. A two chain algorithm of 400 000 iterations with an initial burn-in of 20 000 iterations was run. Convergence was assessed by the Gelman and Rubin diagnostic [8].

OpenBugs code for the CAR model

```
#Childhood diseases associated with U5fever
#CAR Logistic regression model with spatially varying effects of diseases at the regional level

#Likelihood

model {
for (i in 1:M)      #M is the number of observations in the sample
{

had_fever[i]~dbern(p[i])
logit(p[i])<-b0 + part1[i] + part2[i] + part3[i] + part4[i] + part5[i] + part6[i] + part7[i] + part8[i] + part9[i] +
part10[i] + part11[i] + part12[i] + part13[i] + part14[i] + part15[i] + w[clust_id[i]]

part1[i] <- b[1]*savanna_d[i]
part2[i] <- b[2]*lstn[i]
part3[i] <- b[3]*pct_Water[i]
part4[i] <-b[4]*p_soap_water[i]
part5[i] <-b[5]*equals(water[i],1)
part6[i] <-b[6]*equals(cook_fuel[i],1)
part7[i] <-b[7]*equals(married[i],1)
part8[i] <-b[8]*equals(occupat[i],1)
part9[i] <-b[9]*equals(reside[i],1)
part10[i] <-b[10]*equals(wealth_ind[i],1) + b[11]*equals(wealth_ind[i],2) +b[12]*equals(wealth_ind[i],3)
                                          +b[13]*equals(wealth_ind[i],4)
part11[i] <- b[14]*p_rece_bcg[i]
part12[i] <-b[15]*equals(month_cat[i],1) + b[16]*equals(month_cat[i],2) + b[17]*equals(month_cat[i],3)
+                                          b[18]*equals(month_cat[i],4) +
b[19]*equals(month_cat[i],5)
part13[i] <-(b1+ aw1[region[i]])*equals(malaria[i],1)
part14[i] <-(b2+ aw2[region[i]])*equals(all_ari[i],1)
part15[i] <-(b3+ aw3[region[i]])*equals(diarrhoea[i],1)
}

#Priors for betas
b0~dnorm(0,0.01)

for (i in 1:19) {
b[i]~dnorm(0,0.01)
}
```

```
for (i in 1:19) {
or_b[i]<-exp(b[i])
}
```

CAR Spatial covariates for malaria prevalence

```
for(i in 1:sumNumNeigh){
 weights[i]<-1
 }
```

#Malaria

```
b1~dflat()
aw1[1:numRegions]~car.normal(adj[], weights[],num[],tau.car1)
tau.car1 ~ dgamma(2.01,1.01)
sigma.car1 <- sqrt(1/tau.car1)

#ORs of malaria at national and regional level
hrb1<-exp(b1)
for (i in 1:15){
or_aw1[i]<-exp(aw1[i])
}
```

#Symptoms of acute respiratory infections
#all_ari

```
b2~dflat()
aw2[1:numRegions]~car.normal(adj[], weights[],num[],tau.car2)
tau.car2 ~ dgamma(2.01,1.01)
sigma.car2 <- sqrt(1/tau.car2)

#ORs of all_ari at national and regional level
hrb2<-exp(b2)
for (i in 1:15){
or_aw2[i]<-exp(aw2[i])
}
```

#diarrhoea

```
b3~dflat()
aw3[1:numRegions]~car.normal(adj[], weights[],num[],tau.car3)
tau.car3 ~ dgamma(2.01,1.01)
sigma.car3 <- sqrt(1/tau.car3)

#ORs of diarrhoea at national and regional level
hrb3<-exp(b3)
for (i in 1:15){
or_aw3[i]<-exp(aw3[i])
}
```

#Geostatistical random effects at cluster level

```
w[1:N]~spatial.exp(mu[], longnum[], latnum[], tau.sp, phi, 1) #N=number of clusters

for (i in 1:N) {
                                mu[i]<-0
                                }

tau.sp~dgamma(2.01,1.01)
sigma_sp<-1/tau.sp
phi~dunif(0.4558796,312.764)
phi.inv<-1/phi
```

```
Range<-3/phi
#(phi in Winbugs is rho in the notes and tau=1/sigma^2)
}
```

References

[1] George EI, McCulloch RE. Variable selection via Gibbs sampling. Journal of the American Statistical Association 1993;88:881–9. doi:10.2307/2290777.

[2] Giardina F, Kasasa S, Sié A, Utzinger J, Tanner M, Vounatsou P. Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: A spatial and temporal analysis. The Lancet Global Health 2014;2:e601–15. doi:10.1016/S2214-109X(14)70300-6.

[3] Ssempiira J, Nambuusi B, Kissa J, Agaba B, Makumbi F, Kasasa S, et al. Geostatistical modelling of malaria indicator survey data to assess the effects of interventions on the geographical distribution of malaria prevalence in children less than 5 years in Uganda. PLOS ONE 2017;12:e0174948. doi:10.1371/journal.pone.0174948.

[4] Banerjee S, Gelfand A, Carlin B. Hierarchical modeling and analysis for spatial data: Monographs on Statistics and Applied Probability. 2nd ed. New York: Chapman & Hall/CRC; 2015.

[5] Cressie, NAC. Statistics for Spatial Data, Revised Edition: Wiley Series in Probability and Statistics. New York: John Wiley & Sons; 2015.

[6] Bivand RS, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R (Use R!). 2nd ed. New York: Springer, New York, NY; 2013.

[7] Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D. GeoBugs user manual Version 1.2 2004.

[8] Gelman A, Donald R. Inference from iterative simulation using multiple sequences. Statistical Science 1992;7:457–511. doi:10.1214/ss/1177011136.

Table S1. Coverage of interventions at national and regional levels, Uganda DHS 2016

	INT coverage (%)						Vaccination coverage (%)				WASH practices coverage (%)		
Geographical scale	**%hh1itn**	**%hh1itn4two**	**%pp1itn**	**%ppslept**	**%chslept**	**%itnused**	**BCG**	**DPT**	**Polio**	**Measles**	**Water**	**Sanitation**	**Soap/detergent water**
National	78	51	65	55	62	74	96	80	70	80	78	19	44
Regions													
Kampala	75	58	66	60	69	81	99	82	59	83	94	24	71
Central 1	79	59	70	59	67	77	93	76	64	76	70	33	58
Central 2	75	50	65	53	63	73	95	78	59	73	72	31	57
Busoga	75	48	61	52	58	77	97	72	57	70	90	29	22
Bukedi	74	41	54	42	49	69	98	78	62	77	94	15	54
Bugisu	72	39	55	52	60	85	99	73	58	80	84	7	27
Teso	84	48	64	62	72	87	99	92	78	87	95	16	16
Karamoja	55	23	36	33	47	68	99	86	78	91	87	2	15
Lango	79	47	63	53	66	77	96	79	65	75	85	9	23
Acholi	81	41	58	59	68	83	99	85	80	85	81	9	22
West Nile	92	61	77	71	77	73	96	83	74	82	85	4	37
Bunyoro	76	49	62	57	60	77	98	79	75	84	77	15	38
Tooro	77	49	63	50	53	67	96	76	62	87	63	11	57
Ankole	85	58	74	55	58	65	97	84	76	82	53	15	37
Kigezi	89	68	79	55	60	57	98	89	78	96	64	15	31

Table S2. Coverage of treatments and health care seeking at national and regional levels, Uganda DHS 2016

Geographical scale	Treatments coverage (%)			Health care seeking coverage (%)			
	Antibiotics	ORS or RHF	ACT	RDT	Fever advice	ARI advice	Diarrhoea advice
National	29	49	88	49	81	80	71
Regions							
Kampala	46	45	72	55	92	88	71
Central 1	19	53	81	59	88	80	66
Central 2	28	48	90	43	89	85	68
Busoga	37	53	91	43	78	81	72
Bukedi	43	56	89	34	79	81	73
Bugisu	18	40	86	36	91	76	69
Teso	36	31	89	44	64	70	61
Karamoja	26	81	93	68	90	84	85
Lango	20	36	87	49	82	83	86
Acholi	26	55	91	67	85	95	78
West Nile	24	56	90	57	90	93	80
Bunyoro	13	55	90	48	73	93	75
Tooro	16	59	86	57	74	69	65
Ankole	32	30	71	47	80	81	64
Kigezi	21	59	59	37	80	74	71

ARI; symptoms of acute respiratory infections, Antibiotics; Percentage of fever children receiving antibiotics, ORS or RHF; Percentage of children with diarrhea receiving fluid from oral rehydration solution (ORS) sachets or recommended home fluids (RHF), ACT; Percentage of children receiving artemisinin-based combination therapy (ACT) among those with a fever who took any antimalarial drugs (during the 2 weeks period before the survey), Rapid diagnostic test (RDT); Percentage of fever children who had blood taken from a finger or heel for malaria testing, Fever advice; Percentage of fever children for whom advice or treatment was sought from health provider, a health facility, or a pharmacy, ARI advice; Percentage of children with symptoms of ARI for whom advice or treatment was sought from health provider, a health facility, or a pharmacy and Diarrhoea advice; Percentage of children with diarrhea for whom advice or treatment was sought from health provider, a health facility, or a pharmacy.

Table S3. Posterior inclusion probabilities for diseases, interventions, treatments, health care seeking, socio-demographic and environmental/climatic factors

Variable	Inclusion probability (%)	Variable	Inclusion probability (%)
Diseases		**Household factors**	
Malaria	100.0*	Stool disposal	2.0
All ARI	100.0*	**Climatic/environmental factors**	
Diarrhoea	100.0*	**Land cover**	
ITN ownership		% surface covered by forest within a 5km buffer	38.3
None	95.1	% surface covered by water within a 5km buffer	88.5*
%hh1itn	1.0	% surface covered by crop within a 5km buffer	34.0
%hh1itn4two	1.8	Rural or urban	47.8
%pp1itn	2.1	**LST day**	
ITN use		None	100.0
None	92.4	Linear	0.0
%ppslept	0.8	Categorical	0.0
%chslept	0.6	**LST night**	
%itnused	6.2	None	0.0
Vaccinations		Linear	100.0*
BCG	83.0*	Categorical	0.0
DPT	1.0	**NDVI**	
Polio	1.0	None	70.8
Measles	2.0	Linear	29.2
Treatments		Categorical	0.0
ACT	6.0	**Rainfall**	
Antibiotics	1.4	None	91.8
ORS or RHF	1.2	Linear	8.2
WASH practices		Categorical	0.0
Water	98.0*	**Altitude**	
Sanitation	28.0	None	59.0
Soap/detergent and water	93.0*	Linear	41.0
Health care seeking		Categorical	0.0
Rapid diagnostic test	2.0	**Distance to forest**	
Fever advice	2.0	None	98.0
ARI advice	0.0	Linear	2.0
Diarrhoea advice	1.0	Categorical	0.0
Socio-demographic factors		**Distance to water**	
Child factors		None	89.2
Sex	3.4	Linear	10.8
Age of child		Categorical	0.0
None	0.0	**Distance to savanna**	
Linear	0.0	None	49.8
Categorical	100.0*	Linear	50.2*
Residence	88.3*	Categorical	0.0
Maternal factors		**Distance to crops**	
Education level	18.0	None	67.2
Marital status	100.0*	Linear	32.8
Occupation	100.0*	Categorical	0.0
Household factors			
Wealth index	100.0*		
Type of cooking fuel	55.0*		

*Included in the final model with a probability ≥ 50%