Mapping childhood illness: Understanding the distribution of common symptoms in children under five in Africa, 2000 – 2015

Daniel Casey

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2017

Committee:

Simon Hay

Robert Reiner

Program Authorized to Offer Degree:

Global Health

University of Washington

Abstract

Mapping childhood illness: Understanding the distribution of common symptoms in children under five in Africa, 2000 – 2015
Daniel Casey

Chair of the Supervisory Committee:
Professor Simon Hay
Global Health

After twenty years as the World Health Organization's premier strategy for improving childhood health, the Integrated Management of Neonatal and Childhood Illness (IMNCI) strategy is in the midst of an update. Increasing emphasis has been placed on the use of equity and mapping analyses to target implementation and adjust programs to local epidemiology. Our study interacts with these policy movements by producing geostatistical maps of the period prevalence of common symptoms of the childhood illnesses identified by the IMNCI guidelines. Specifically, we use Bayesian model-based geostatistics to create yearly maps of cough, diarrhea, fever, and their comorbidity in Africa from 2000 – 2015 at a 5x5 km^2 resolution. These maps reveal uneven progress in the reduction of prevalence of these symptoms since the new millennium-- both between and within countries. The maps highlight areas where intervention should be focused and provide a high-level description of the underlying epidemiology. We expect that these geospatial products will allow policymakers to deploy interventions that are spatially and epidemiologically targeted to reduce the burden of childhood illness.

Contents

Introduction

During the 1990s, the World Health Organization (WHO) and UNICEF developed and began deploying the Integrated Management of Neonatal and Childhood Illness (IMNCI) strategy for reducing child mortality, with an emphasis on mortality due to acute respiratory infection, diarrhea, measles, malaria and malnutrition.[1–3] The strategy features three main components: (1) effective and integrated case management of childhood illness, (2) health system strengthening and (3) familial and community outreach and education.[3]

In the ensuing 20 years, IMNCI has been rolled out broadly and evidence suggests that the strategy is effective at reducing childhood mortality.[1,4,5] With the Sustainable Development Goals (SDG) looming, and in light of its history of variable implementation and uncertain funding, there have been calls for the revision and revitalization of the IMNCI strategy.[1] The resulting efforts suggest modifying the strategy to provide a greater emphasis on reaching underserved populations and adjusting protocols to match local epidemiology as informed by mapping exercises.[1]

The availability of spatially referenced data and advancements in statistical computing have catalyzed mapping efforts within the field of spatial epidemiology. In the last decade, maps of IMNCI-covered diseases like malaria have provided policymakers and health workers with a substantially improved level of spatial precision.[6,7] However, the spatial and temporal coverage of existing maps vary from disease to disease, and little work has been done to directly examine the comorbidity or interaction of these diseases. Where scholars have examined the comorbidity of IMNCI conditions, spatial resolution is often at the country or regional level.[8] Since a crucial component of the IMNCI strategy is holistic case management, work examining how such illnesses may be correlated in space to help assess patterns of comorbidity is needed. These insights can subsequently be used to plan and evaluate joint interventions.

The present study begins to address the spatial and epidemiologic gaps related to the IMNCI strategy via high-resolution (5km x 5km) gridded maps of common childhood symptoms. Specifically, we examined the relative distribution of cough, diarrhea, fever, and their comorbidity in Africa from 2000 to 2015. This analysis emphasized symptoms rather than diseases for a few reasons. First, cough, diarrhea and fever feature prominently in the ICMNI algorithm for childhood illness case management and otherwise constitute the primary symptoms of the main contributors to childhood mortality.[9,10] Second, population representative household surveys from series like the Demographic and Health Survey and the Multiple Indicator Cluster Survey routinely ask about symptoms rather than disease, given the low specificity of disease status from recall.[11] Third, information on the comorbidity of common childhood diseases is relatively sparse.

Methods

Overview

This analysis provides estimates of the period prevalence of fever, cough, diarrhea and their comorbidity in children under five years old in Africa from 2000 – 2015 at the 5km^2 level. Specifically, we leveraged a large database of geopositioned household surveys and Bayesian model-based geostatistics (MBG) to make maps of eleven different variables from a system of four main equations:

Eq 1.: F = c'd'f + c'df + cd'f + cdf
Eq 2.: D = c'df' + c'df + cdf' + cdf
Eq 3.: C = cd'f' + cd'f + cdf' + cdf
Eq 4.: c'd'f' = 1 – (cd'f' + cd'f + cdf' + cdf + c'df' + c'd'f + c'df)

C, D and F refer to cough, diarrhea, and fever, respectively. For notational clarity and consistent diction, these symptom aggregates are termed "symptom groups" while the intersections (e.g. the components on the right hand side of equations 1 – 3) are termed "event categories." These are represented with lower case letters while the apostrophe marks a symptom "not" in the group. For example, c'd'f can be interpreted as fever without cough or diarrhea. Figure 1 visualizes the relationship between the estimated elements via a Venn diagram. We use the term "indicator" to describe event categories and symptom groups generically.

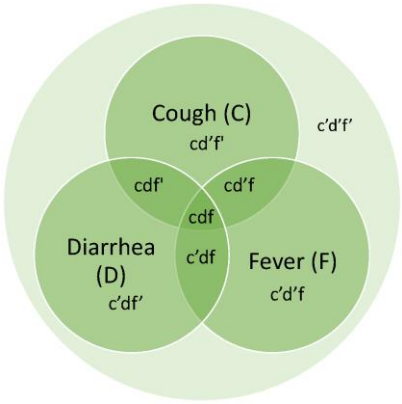

Figure 1: Venn diagram visualization of the comorbidity of childhood illness

Using an iterative reconciliation procedure, we searched for a solution to the series of equations detailed above. Details on the data and methods are provided below with additional explanation in the appendix.

Survey Data

We extracted data symptom status of 1.36 million children from 182 different household surveys from 47 countries. Symptom status was derived from questions with the general form: "Has child X had symptom Y in the last two weeks?", where Y is one of fever, cough or diarrhea. We then

categorized each child into one or more of the aforementioned categories depending on their (mother's) response to this series of questions. If a survey did not capture information about all three symptoms, children were not classified to an event category. These children were classified, where possible, to a symptom group. Of the surveys included in this study, only 21 did not have information on all three symptoms. These are primarily Malaria Indicator Surveys that did not ask about diarrhea. A full listing of the surveys used for this analysis is available in the appendix.

After classification, data from surveys with geo-referenced cluster coordinates were aggregated to the cluster level assuming a simple random sample. In surveys where cluster level spatial positioning was not available, we aggregated the microdata to the smallest available administrative unit boundaries using the survey package in R.[12] These boundary/polygon level results were then resampled into disperse points as a function of the underlying spatial distribution of population.[13]

Covariates and Stacked Generalization

We assembled 26 spatially and temporally explicit covariate layers at a $5km^2$ resolution with an additional two country-year level covariates. These covariates included access to major population centers, measures of aridity, greenness and other environmental metrics derived from MODIS, climatic metrics like diurnal range, and distance from water features. A full list and appropriate is provided in the appendix. The country-year level covariates were the sociodemographic index introduced by the Global Burden of Disease project and the health access and quality index.[10,14] The latter two covariates were included to capture social determinants of health not currently available at a finer spatial resolution. For each data point, we extracted the relevant covariate values by space and time to train our models.

To capture nonlinear and interactive effects between our covariates and the outcome, we used stacked generalization as a preprocessing step.[15] Stacked generalization is a method of model ensembling whereby a number of different "child" models/algorithms are trained and return cross-validated predictions and in-sample predictions. A parent model is subsequently fit on the cross-validated predictions of the child models while the in-sample predictions from the child models are used to generate predictions from the parent model. This exchange is done to reduce overfitting in the parent model while utilizing all the information available. For this analysis, we used generalized additive models, boosted regression trees, and penalized regression with an elastic net penalty as the child models.[16–18] For the parent model, we used the MBG model described in the next step.

Analysis

We fit 55 separate MBG models for this analysis— one for every indicator and region pair. The MBG model is a Gaussian Process Regression (GPR) fit on the results of the stacked generalization process with the residuals modeled through three separate random effect terms. First, we employed a "nugget" random effect that captures data source specific variation. Second, we used a random effect on country to capture country-specific deviation and attenuate

covariate effects. The final random effect is the Kronecker product of a temporal autoregressive-one term and an approximation of a Gaussian Markov random field via stochastic partial differential equations. This last term modelled spatiotemporal correlation between adjoining years as well as between nearby areas. From each model fit, we generated 100 draws from the posterior distribution to create a series of candidate maps. All MBG models were fit using R-INLA.[19,20]

Reconciliation

Since we independently estimated the components of a system of equations, we enforced the underlining constraints post-hoc. We iterated through each equation and attempted to find the multiplicative scalar that would solve the equation. For example, when we solve equation 4, we would find a value X such that the left hand side of the equation equals X times the sum of the right hand side. We subsequently apply a fraction of the implied change (e.g. X * indicator - indicator) to each variable, except for one held constant, where the movement towards the adjusted value is moderated by a learning rate and the relative uncertainty of the variable. We defined the relative uncertainty for a given indicator as:

$$(1_{indicator}/N_{indicator})/\sum_{j=1}^{11}(1_{indicator\ j}/N_{indicator\ j})$$

Where N is back-calculated from the equation of a binomial standard deviation per pixel and *j* is the index of the indicator.

Tabulation and Visualization

We used conditional simulation to generate summary measures at district and country levels.[21] For ease of visualization, we masked our maps to exclude areas with less than 10 people per $5km^2$.

Results

Continent-wide period prevalence of fever, cough, diarrhea and their comorbidity

Table 1 shows event category results for Africa for the years 2000, 2005, 2010, and 2015. In 2015 in Africa, 64% of children were free of cough, diarrhea and fever. Of the remaining children, 8% had cough only, 6.7% had fever only, 5% had diarrhea only, 6.5% had cough and fever without diarrhea, 2.8% had cough and diarrhea without fever, 2.8% had fever and diarrhea without cough, and a final 3.4% had all three symptoms. Of the major symptom groups, cough was the most prevalent at 21%, while 20% of children were estimated to have fever and 14% with diarrhea. Since 2000, the estimated period prevalence of fever and cough both dropped about 10 percentage points while diarrhea dropped 5 percentage points.

Table 1: Period prevalence (in %) of the eight event categories for Africa.

Year	*cdf'*	*cd'f*	*cdf*	*cd'f'*	*c'df*	*c'df'*	*c'd'f*	*c'd'f'*
2000	3.8	9.9	6.3	11.3	4.3	5.6	9.6	49.2
2005	2.8	8.2	4.8	9.2	2.9	5.4	6.9	59.9
2010	3.6	7.1	3.9	12.4	2.8	5.6	6.4	58.1

2015	2.8	6.6	3.5	8.2	2.8	5.3	6.8	64.1

Figure 2 shows a panel of maps of the three symptom groups for the year 2015. Of the three symptom groups, cough has the most spatially variable distribution with hotspots of about 30-50% prevalence in the north of the Democratic Republic of Congo, eastern Zimbabwe, a swath along the southern border of Kenya, and the border region of Zambia, Zimbabwe, Botswana, Angola and Namibia. Areas with lower prevalence of cough (less than 10%) include northern Kenya, much of western Sub-Saharan Africa, and northern Zambia. For diarrhea, estimated period prevalence is relatively low and constant over space compared to cough and fever. Swaziland, Northern Nigeria, and the southern parts of Sudan have amongst the highest levels of diarrhea. Finally, for fever, the central part of the continent, in particular the Democratic Republic of Congo (DRC), and Central African Republic have high prevalence. In the Lake Victoria area, we estimate the prevalence of children with fever is above 30% in most of the adjacent districts.

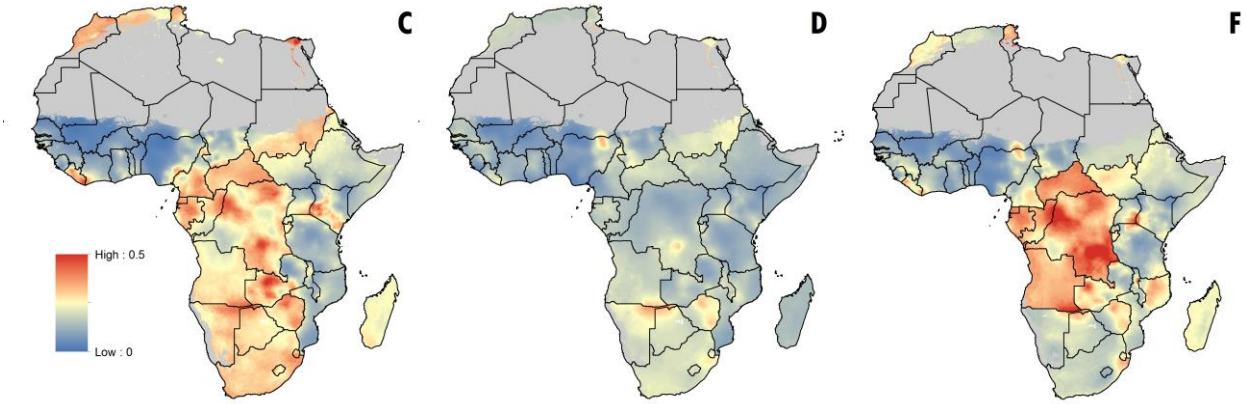

Figure 2: Period Prevalence of the main symptom groups for 2015. Areas with low/no population are masked in gray. Recall that C represents the period prevalence of with cough (regardless of other symptoms), while D and F refer to diarrhea and fever, respectively.

Change over time in the prevalence of fever, cough and diarrhea in Africa

Figure 3 provides a panel of maps showing the annualized rate of change between 2000 and 2015 for the 3 symptom groups and children without illness. These maps show a general decrease in the period prevalence of all three symptom groups in the last 15 years. Decrease in the prevalence of any symptoms is noticeable in the western Sub-Saharan Africa region—specifically in countries like Ghana, Ivory Coast, and Nigeria. Other locations, like Tanzania and southern Mozambique have similar patterns. Of the three main symptom groups, cough has experienced the broadest geographic decline, with fever a close second. Despite the general decline in childhood symptoms, there are some noticeable areas with stagnant or negative progress. The middle and northern parts of Cameroon, southern DRC, and the northern districts in the Republic of Congo in particular have not made the same improvements since 2000 as their regional and subnational neighbors.

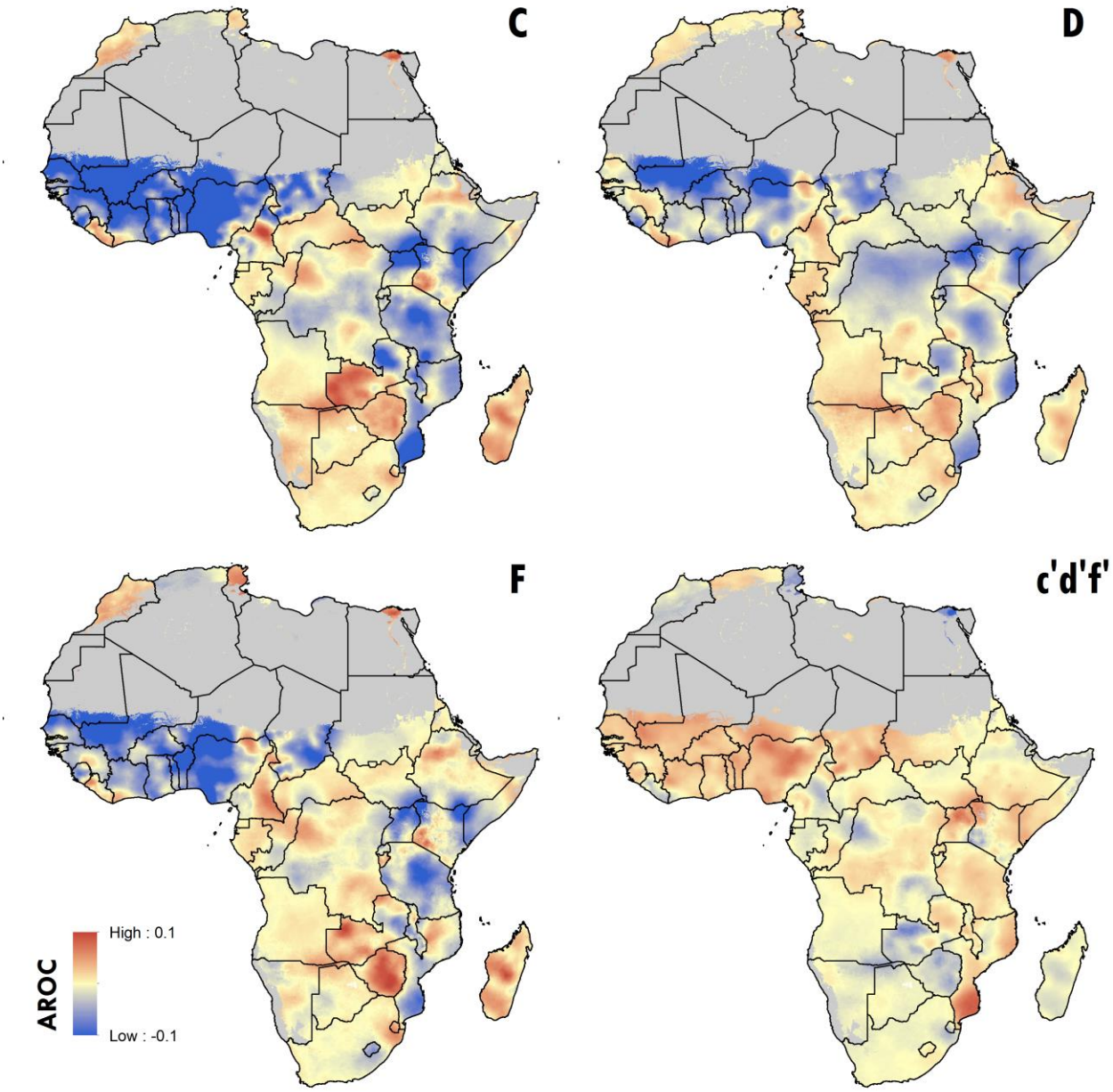

Figure 3: Annualized rate of change (AROC) from 2000-2015 for the main symptom groups and no symptoms. The method of calculating AROC presented here calculates the pairwise change between years while weighting years closer to 2015 more heavily.

Understanding the spatial distribution of fever and its components

The spatial heterogeneity observed for 2015 and for the entire time series between the symptom groups is also evident within the three main symptom groups. Figure 4 provides a panel of maps representing the mean period prevalence of each event category that makes up fever for 2015. Perhaps unsurprisingly, the event categories representing the overlap between fever and diarrhea are visibly lower than fever without cough or diarrhea, and cough and fever without diarrhea. These latter event categories likely represent infectious diseases without localizing features (e.g. malaria or dengue) and lower respiratory infection respectively. While fever without cough or

diarrhea is concentrated in southwest DRC and northern Mozambique, fever and cough without diarrhea has relatively high prevalence in the north of the DRC, spilling over into the Republic of Congo and the Central African Republic.

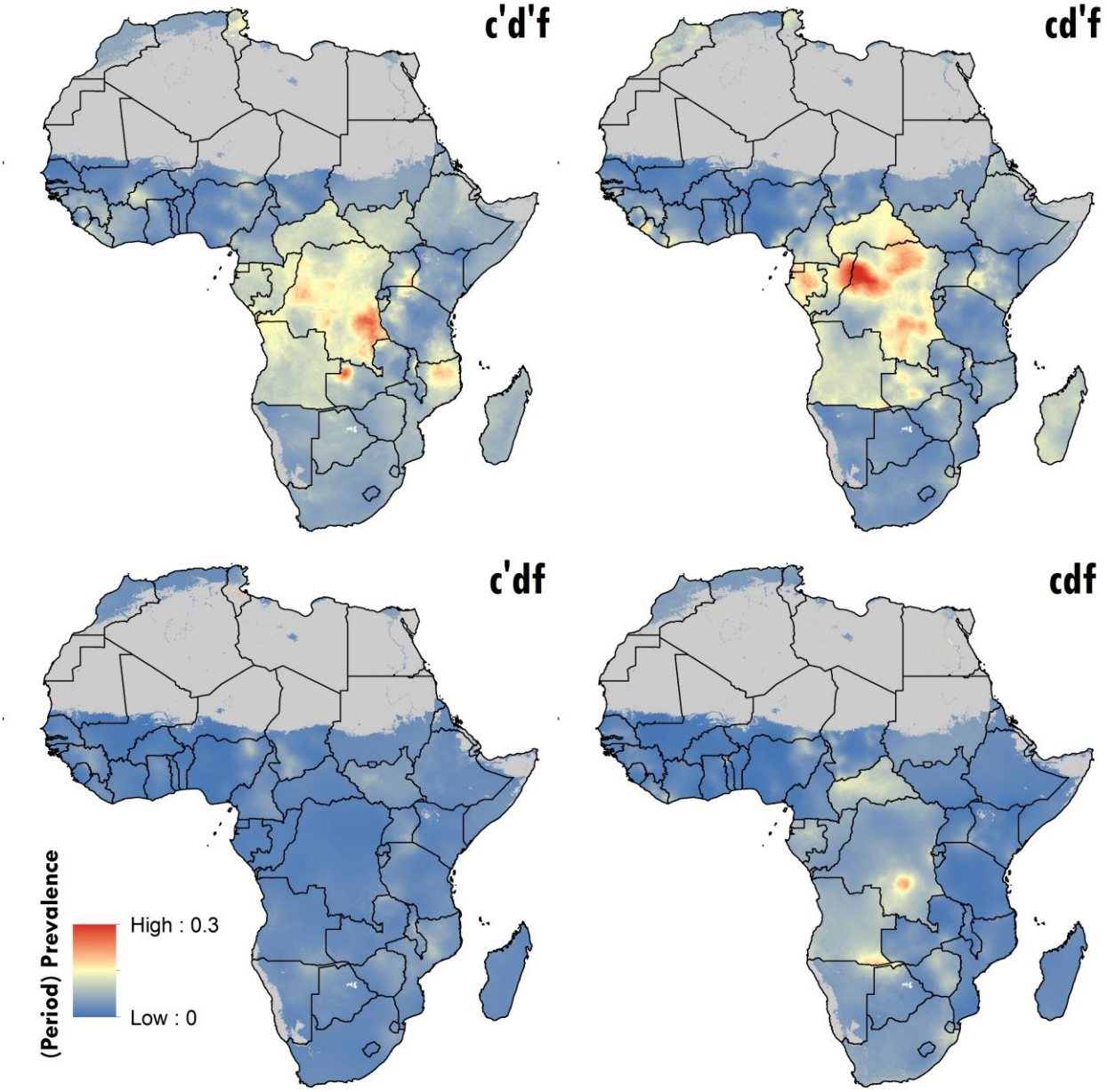

Figure 4: Fever and its comorbidity/co-presentation with other common symptoms. Africa, 2015.

Discussion

The role of mapping in the future of the IMNCI strategy

In late 2016, an international group of experts specified five problem areas in the IMNCI strategy implementation in need of amelioration: (1) fragmentation of global strategies related to

childhood health, (2) lack of adequate funding and need to target delivery, (3) lack of systematic evidence of impact and effectiveness, (4) lack of tailored solutions based on epidemiology and health systems and (5) lack of accountability and clear targets.[1] In this study, we provide the first set of high-resolution maps of three common symptoms of childhood illness and their comorbidity for Africa -- a timely resource as the IMNCI guidelines are expanded, adapted, improved, and implemented.

With the prioritization of the Sustainable Development Goals (SDG), geographic and demographic equity are increasingly a point of emphasis. The maps presented here provide a historical perspective on the equity of health gains of the last fifteen years and highlight areas in need of improvement. For example, estimates suggest that the more prosperous areas of Zambia, like Lusaka, have had a sustained reduction in the measured symptoms over the last 15 years. By contrast, the western part of the country remains relatively behind. The maps reveal similar storylines in other countries like Nigeria, Kenya, and Cameroon, where there are visible subnational differences in the decline of childhood illness.

High spatial resolution estimates of the current situation, in conjunction with an understanding of an area's epidemiological history, provide crucial pillars for resource allocation decisions and are integral to an equity-based approach. Furthermore, these maps serve as a benchmark by which future progress can be measured. By adopting the same holistic outlook promoted by the IMNCI guidelines, policymakers can and should evaluate their efforts with these data. For example, the northeast part of Mozambique has seen a decline in all three symptoms over the last 15 years. However, in other parts of the country, only one or two of the indicators has an observable decline. The former situation potentially implies efforts to strengthen health care systems, while symptom-specific declines could highlight disease-specific vertical programming. Such theses must be matched with local knowledge, but at the very least, these maps provide the opportunity to ask these important questions. Moreover, as a snapshot of the (nearly current) situation, policymakers can set reduction targets using these estimates as a baseline.

Limitations and future work

The estimates provided in this paper as a useful first step towards providing spatial data to inform the future of the IMNCI guidelines. However, these maps are not static products, and to remain relevant they must be updated as new data becomes available. While the present results are based off a large collection of nationally representative household surveys, other usable forms of data surely exist. The inclusion of administrative reports and other spatially fine data will improve the certainty current results, highlight new places where success has been achieved, and identify areas where improvement is needed. Moreover, the process by which these data are identified, collated, cleaned, shared, and analyzed is aligned with and supportive of the stated goals of the IMNCI guidelines.

Methodologically, there are a few clear paths for future study. First, the post-hoc enforcement of the system of equations through iterative raking may not respect the underlying error correlation and has the potential to break the spatial field underlying the analysis. The raking process

requires significant computing resources making it difficult to be used more widely. This concern may dissipate over time as model-based geostatistics gains the capability to produce multinomial models at scale, where the estimates will a-priori be internally consistent. Second, a shift away from period prevalence towards incidence as the measure of interest will allow for clearer explanations and otherwise link up more cleanly with common ways of describing illnesses with short duration. Third, expanding the study area to include other countries covered by the IMNCI strategy would correspondingly increase the usefulness and audience of these maps. Fourth, the translation of symptoms into illness and the expansion of the maps to include other diseases emphasized by the IMNCI will provide additional epidemiological granularity. Splitting out the main symptoms by disease and etiology will allow for an easier link up with other health metrics efforts like the Global Burden of Disease project.[22] For example, the fraction of fevers due to malaria has declined in the last fifteen years, highlighting the need for effective diagnostics, treatment and understanding of the remaining causes of fever.[23]

Conclusion

As the third decade of the IMNCI strategy takes shape and the SDG era begins in earnest, so too begins a renewed push to improve childhood health globally. Despite marked progress in the last decade and a half, Africa still bears a large burden of disease from the childhood illnesses addressed by the IMNCI strategy.[10] In this new era, holistic, equitable, and evidence-based interventions are the primary way by which progress will be made. The maps presented in this paper support this next wave of interventions. By providing spatial granularity, we reveal contemporary continental, regional, national, and subnational disparities in childhood health. When viewed over time these results allow policymakers to learn from areas where progress has been achieved previously and apply those lessons to new areas.

Citations

1 Costello A, Dalglish S, Aboubaker S, *et al.* Towards a Grand Convergence for child survival and health: A strategic review of options for the future building on lessons learnt from IMNCI. *Adolesc Heal Heal Res Dev Team* 2016. http://www.who.int/iris/bitstream/10665/251855/1/WHO-MCA-16.04-eng.pdf?ua=1.

2 Costello A. Integrated management of childhood illness. *Lancet (London, England)* 1997; **350**: 1266.

3 Tulloch J. Integrated approach to child health in developing countries. *Lancet (London, England)* 1999; **354 Suppl**: SII16-I20.

4 Gera T, Shah D, Garner P, Richardson M, Hs S. Integrated management of childhood illness (IMCI) strategy for children under five (Review) SUMMARY OF FINDINGS FOR THE MAIN COMPARISON. 2016. DOI:10.1002/14651858.CD010123.pub2.www.cochranelibrary.com.

5 Schellenberg JRA, Adam T, Mshinda H, *et al.* Effectiveness and cost of facility-based Integrated Management of Childhood Illness (IMCI) in Tanzania. *Lancet* 2004; **364**: 1583–94.

6 Gething PW, Casey DC, Weiss DJ, *et al.* Mapping *Plasmodium falciparum* Mortality in Africa between 1990 and 2015. *N Engl J Med* 2016; **375**: 2435–45.

7 Bhatt S, Weiss DJ, Cameron E, *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* 2015; **526**: 207–11.

8 Prasad N, Sharples KJ, Murdoch DR, Crump JA. Community prevalence of fever and relationship with malaria among infants and children in low-resource areas. *Am J Trop Med Hyg* 2015; **93**: 178–80.

9 WHO IMCI. Integrated Management of Childhood Illness (IMCI) Chart Booklet. *Distance Learn Course* 2014; : 1–76.

10 Wang H, Coates MM, Coggeshall M, *et al.* Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1725–74.

11 Geldsetzer P, Williams TC, Kirolos A, *et al.* The recognition of and care seeking behaviour for childhood illness in developing countries: A systematic review. *PLoS One* 2014; **9**. DOI:10.1371/journal.pone.0093427.

12 Lumley T. Survey: analysis of complex survey samples. 2017. https://cran.r-project.org/web/packages/survey/.

13 Tatem AJ, Garcia AJ, Snow RW, *et al.* Millennium development health metrics: where do Africa's children and women of childbearing age live? *Popul Health Metr* 2013; **11**: 11.

14 Collaboraters G 2015 HA and Q. Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990-2015: A novel analysis from the Global Burden of Disease Study 2015. *Lancet* 2017; : 1990–2015.

15 Bhatt S, Cameron E, Flaxman SR, Weiss DJ, Smith DL, Gething PW. Improved prediction accuracy for disease risk mapping using Gaussian Process stacked generalisation. 2016; : 1–33.

16 Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Ser B Stat Methodol* 2011; **73**: 3–36.

17 Chen T, Guestrin C. XGBoost : Reliable Large-scale Tree Boosting System. *arXiv* 2016; : 1–6.
18 Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010; **33**. https://www.jstatsoft.org/article/view/v033i01.
19 Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol* 2009; **71**: 319–92.
20 Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc B* 2011; **73**: 423–98.
21 Gething PW, Patil AP, Hay SI. Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Comput Biol* 2010; **6**. DOI:10.1371/journal.pcbi.1000724.
22 GBD. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infect Dis* 2017; **3099**: 1–40.
23 Prasad N, Murdoch DR, Reyburn H, Crump JA. Etiology of severe febrile illness in low- and middle-income countries: A systematic review. *PLoS One* 2015; **10**: 1–25.

Appendix

Table of Survey Data

Survey Name	ISO3 Code	Survey Start Year	Survey End Year
MACRO AIS	TZA	2007	2008
MACRO AIS	TZA	2011	2012
MACRO DHS	BFA	1998	1999
MACRO DHS	CIV	1998	1999
MACRO DHS	CMR	1998	1998
MACRO DHS	GHA	1998	1999
MACRO DHS	KEN	1998	1998
MACRO DHS	NER	1998	1998
MACRO DHS	TGO	1998	1998
MACRO DHS	ZAF	1998	1998
MACRO DHS	GIN	1999	1999
MACRO DHS	TZA	1999	1999
MACRO DHS	ZWE	1999	1999
MACRO DHS	EGY	2000	2000
MACRO DHS	ETH	2000	2000
MACRO DHS	MWI	2000	2000
MACRO DHS	NAM	2000	2000
MACRO DHS	RWA	2000	2000
MACRO DHS	UGA	2000	2001
MACRO DHS	BEN	2001	2001
MACRO DHS	ZMB	2001	2002
MACRO DHS	BFA	2003	2003
MACRO DHS	EGY	2003	2003
MACRO DHS	GHA	2003	2003
MACRO DHS	KEN	2003	2003
MACRO DHS	MAR	2003	2004
MACRO DHS	MDG	2003	2004
MACRO DHS	MOZ	2003	2003
MACRO DHS	NGA	2003	2003
MACRO DHS	CMR	2004	2004
MACRO DHS	LSO	2004	2005
MACRO DHS	MWI	2004	2005
MACRO DHS	TCD	2004	2004
MACRO DHS	TZA	2004	2005
MACRO DHS	COG	2005	2005
MACRO DHS	EGY	2005	2005
MACRO DHS	ETH	2005	2005

MACRO DHS	GIN	2005	2005
MACRO DHS	RWA	2005	2005
MACRO DHS	SEN	2005	2005
MACRO DHS	ZWE	2005	2006
MACRO DHS	BEN	2006	2006
MACRO DHS	LBR	2006	2007
MACRO DHS	MLI	2006	2006
MACRO DHS	NAM	2006	2007
MACRO DHS	NER	2006	2006
MACRO DHS	SWZ	2006	2007
MACRO DHS	UGA	2006	2006
MACRO DHS	COD	2007	2007
MACRO DHS	RWA	2007	2008
MACRO DHS	ZMB	2007	2007
MACRO DHS	EGY	2008	2008
MACRO DHS	GHA	2008	2008
MACRO DHS	KEN	2008	2009
MACRO DHS	MDG	2008	2009
MACRO DHS	NGA	2008	2008
MACRO DHS	SLE	2008	2008
MACRO DHS	STP	2008	2009
MACRO DHS	LSO	2009	2010
MACRO DHS	TZA	2009	2010
MACRO DHS	BDI	2010	2011
MACRO DHS	BFA	2010	2011
MACRO DHS	ETH	2010	2011
MACRO DHS	MWI	2010	2010
MACRO DHS	RWA	2010	2011
MACRO DHS	SEN	2010	2011
MACRO DHS	ZWE	2010	2011
MACRO DHS	BEN	2011	2012
MACRO DHS	CIV	2011	2012
MACRO DHS	CMR	2011	2011
MACRO DHS	COG	2011	2012
MACRO DHS	MOZ	2011	2011
MACRO DHS	UGA	2011	2011
MACRO DHS	COM	2012	2013
MACRO DHS	GAB	2012	2012
MACRO DHS	GIN	2012	2012
MACRO DHS	MLI	2012	2013
MACRO DHS	SEN	2012	2013

MACRO DHS	COD	2013	2013
MACRO DHS	GMB	2013	2013
MACRO DHS	LBR	2013	2013
MACRO DHS	NAM	2013	2013
MACRO DHS	NGA	2013	2013
MACRO DHS	SLE	2013	2013
MACRO DHS	TGO	2013	2013
MACRO DHS	ZMB	2013	2014
MACRO DHS	EGY	2014	2014
MACRO DHS	GHA	2014	2014
MACRO DHS	KEN	2014	2014
MACRO DHS	LSO	2014	2014
MACRO DHS	RWA	2014	2015
MACRO DHS	SEN	2014	2014
MACRO DHS	TCD	2014	2015
MACRO DHS	MWI	2015	2016
MACRO DHS	SEN	2015	2015
MACRO DHS	TZA	2015	2016
MACRO DHS	ZWE	2015	2015
MACRO MIS	SEN	2006	2006
MACRO MIS	LBR	2008	2009
MACRO MIS	SEN	2009	2009
MACRO MIS	UGA	2009	2010
MACRO MIS	NGA	2010	2010
MACRO MIS	AGO	2011	2011
MACRO MIS	LBR	2011	2011
MACRO MIS	MDG	2011	2011
MACRO MIS	BDI	2012	2013
MACRO MIS	MWI	2012	2012
MACRO MIS	RWA	2012	2013
MACRO MIS	MDG	2013	2013
MACRO MIS	BFA	2014	2014
MACRO MIS	MWI	2014	2014
MACRO MIS	UGA	2014	2015
MACRO MIS	KEN	2015	2015
MACRO MIS	MLI	2015	2015
UNICEF MICS	ZMB	1999	1999
UNICEF MICS	BDI	2000	2000
UNICEF MICS	CAF	2000	2000
UNICEF MICS	CIV	2000	2000
UNICEF MICS	CMR	2000	2000

UNICEF MICS	COM	2000	2000
UNICEF MICS	GMB	2000	2000
UNICEF MICS	GNB	2000	2000
UNICEF MICS	GNQ	2000	2000
UNICEF MICS	KEN	2000	2000
UNICEF MICS	MDG	2000	2000
UNICEF MICS	NER	2000	2000
UNICEF MICS	RWA	2000	2000
UNICEF MICS	SDN	2000	2000
UNICEF MICS	SEN	2000	2000
UNICEF MICS	SLE	2000	2000
UNICEF MICS	SSD	2000	2000
UNICEF MICS	STP	2000	2000
UNICEF MICS	SWZ	2000	2000
UNICEF MICS	TCD	2000	2000
UNICEF MICS	TGO	2000	2000
UNICEF MICS	AGO	2001	2001
UNICEF MICS	COD	2001	2001
UNICEF MICS	BDI	2005	2005
UNICEF MICS	GMB	2005	2006
UNICEF MICS	SLE	2005	2005
UNICEF MICS	BFA	2006	2006
UNICEF MICS	CAF	2006	2006
UNICEF MICS	CIV	2006	2006
UNICEF MICS	CMR	2006	2006
UNICEF MICS	DJI	2006	2006
UNICEF MICS	GHA	2006	2006
UNICEF MICS	GNB	2006	2006
UNICEF MICS	MWI	2006	2006
UNICEF MICS	SOM	2006	2006
UNICEF MICS	TGO	2006	2006
UNICEF MICS	GHA	2007	2008
UNICEF MICS	KEN	2007	2007
UNICEF MICS	MRT	2007	2007
UNICEF MICS	NGA	2007	2007
UNICEF MICS	KEN	2008	2008
UNICEF MICS	MOZ	2008	2009
UNICEF MICS	ZWE	2009	2009
UNICEF MICS	CAF	2010	2011
UNICEF MICS	COD	2010	2010
UNICEF MICS	GHA	2010	2011

UNICEF MICS	SDN	2010	2010
UNICEF MICS	SLE	2010	2010
UNICEF MICS	SSD	2010	2010
UNICEF MICS	SWZ	2010	2010
UNICEF MICS	TCD	2010	2010
UNICEF MICS	TGO	2010	2010
UNICEF MICS	GHA	2011	2011
UNICEF MICS	MRT	2011	2011
UNICEF MICS	NGA	2011	2011
UNICEF MICS	SOM	2011	2011
UNICEF MICS	SOM	2011	2011
UNICEF MICS	EGY	2013	2014
UNICEF MICS	MWI	2013	2014
UNICEF MICS	GNB	2014	2014
UNICEF MICS	STP	2014	2014
UNICEF MICS	SWZ	2014	2014
UNICEF MICS	ZWE	2014	2014
UNICEF MICS	MLI	2009	2010
UNICEF MICS	CMR	2014	2014
MACRO DHS	ERI	2002	2002
MACRO DHS	GAB	2000	2001
MACRO DHS	MRT	2000	2001

Table of Spatial Covariates

Covariate	Temporal resolution	Source	Reference
Aridity (via two calculations)	Annual	Climatic Research Unit Time-Series (CRUTS)	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).
Average daily maximum temperature	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at:

Covariate	Temporal resolution	Source	Reference
			https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).
Average Land Surface Temperature (LST)	Annual	MODIS	USGS & NASA. Land surface temperature and emissivity 8-day L3 global 1km MOD11A2 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2. (Accessed: 24th July 2017) Wan, Z. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens*. **98**, 106–118 (2014).
Daytime LST	Annual	MODIS	USGS & NASA. Land surface temperature and emissivity 8-day L3 global 1km MOD11A2 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2. (Accessed: 24th July 2017) Wan, Z. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens*. **98**, 106–118 (2014).
Distance to rivers (and lakes)	Static	Natural Earth Data (derived)	Natural Earth. Rivers and lake centerlines dataset. Available at: http://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-rivers-lake-centerlines/. (Accessed: 24th July 2017)
Diurnal difference in LST	Annual	MODIS	USGS & NASA. Land surface temperature and emissivity 8-day L3 global 1km MOD11A2 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2. (Accessed: 24th July 2017)

Covariate	Temporal resolution	Source	Reference
			Wan, Z. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens*. **98**, 106–118 (2014).
Diurnal temperature range	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol*. **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).
Enhanced Vegetation Index (EVI)	Annual	MODIS	Huete, A., Justice, C. & van Leeuwen, W. MODIS vegetation index (MOD 13) algorithm theoretical basis document. (1999). USGS & NASA. Vegetation indices 16-Day L3 global 500m MOD13A1 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13a1. (Accessed: 25th July 2017) Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens*. **98**, 106–118 (2014).
Frost day frequency	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol*. **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017)
Growing season length	Static	FAO	FAO. GAEZ - Global Agro-Ecological Zones data portal. Available at:

Covariate	Temporal resolution	Source	Reference
			http://www.fao.org/nr/gaez/about-data-portal/en/. (Accessed: 25th July 2017) FAO. GAEZ - Global Agro-Ecological Zones users guide. (2012).
Irrigation	Static	University of Frankfurt	Goethe-Universität. Generation of a digital global map of irrigation areas. Available at: https://www.uni-frankfurt.de/45218039/Global_Irrigation_Map. (Accessed: 25th July 2017)
Malaria incidence	Annual	Malaria Atlas Project	Bhatt, S. et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).
Nighttime LST	Annual	MODIS	USGS & NASA. Land surface temperature and emissivity 8-day L3 global 1km MOD11A2 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2. (Accessed: 24th July 2017) Wan, Z. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD). Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens*. **98**, 106–118 (2014).
Nighttime lights	Annual	NOAA DMSP	NOAA. Version 4 DMSP-OLS nighttime lights time series dataset. Available at: https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. (Accessed: 25th July 2017)
Normalized Difference Vegetation Index (NDVI)	Annual	AVHRR	NASA & NOAA. Advanced Very High Resolution Radiometer (AVHRR) Normalized Difference Vegetation Index (NDVI) dataset. Available at: https://nex.nasa.gov/nex/projects/1349/. (Accessed: 25th July 2017)
Population	Annual	WorldPop	Lloyd, C. T., Sorichetta, A. & Tatem, A. J. High resolution global gridded data for use in population studies. *Sci. Data* **4**, sdata20171 (2017).

Covariate	Temporal resolution	Source	Reference
			World Pop. Get data. Available at: http://www.worldpop.org.uk/data/get_data/. (Accessed: 25th July 2017)
Potential Evapotranspiration (PET)	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).
Precipitation	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).
Tassled cap brightness	Annual	MODIS	USGS & NASA. Nadir BRDF- Adjusted Reflectance Reflectance 16-Day L3 Global 1km dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd43b4. (Accessed: 25th July 2017) Strahler, A. H. & Muller, J.-P. MODIS BRDF/Albedo product: algorithm theoretical basis document version 5.0. (1999). Weiss, D. J. et al. An effective approach for gap-filling continental scale remotely sensed time-series. *Isprs J. Photogramm. Remote Sens.* **98**, 106–118 (2014).
Tassled cap wetness	Annual	MODIS	USGS & NASA. Nadir BRDF- Adjusted Reflectance Reflectance 16-Day L3 Global 1km dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd43b4. (Accessed: 25th July 2017)

Covariate	Temporal resolution	Source	Reference
			Strahler, A. H. & Muller, J.-P. MODIS BRDF/Albedo product: algorithm theoretical basis document version 5.0. (1999).
Travel time to nearest settlement >50,000 inhabitants	Static	European Commission Joint Research Center	European Commission Joint Research Centre. Travel time to major cities: A global map of Accessibility. Available at: http://forobs.jrc.ec.europa.eu/products/gam/index.php. (Accessed: 25th July 2017) Uchida, H. & Nelson, A. Agglomeration index: towards a new measure of urban concentration. (2008).
Urbanicity	Annual	European Commission/ GHS	Pesaresi, M. et al. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. (Publications Office of the European Union, 2016).
Wet day frequency	Annual	CRUTS	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol*. **34**, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/. (Accessed: 24th July 2017).

Data preparation

Microdata were tabulated (aggregated/summarized) at the finest available spatial resolution. For approximately half of the surveys, this was the cluster/primary sampling unit—generally a village or geographically clustered group of households. Tabulations at the cluster level assumed a simple random sample and that the area of the cluster was less than 5x5 km^2. These data were subsequently assigned a latitude and longitude at the centroid. Each "true" cluster was given an analytical weight of 1.

For the remaining surveys, observations were georeferenced and tabulated to the smallest possible administrative (polygon) unit and subsequently transformed into a collection of pseudoclusters. First, for each polygon-level aggregate, we sampled 10,000 points proportional to the underlying spatial distribution of population. Second, the sample points were then clustered via K-means with a target of 5 psuedoclusters per 1,000 pixels in the polygon. Finally, analytical weights were assigned based of the relative proportion of sample points grouped into a pseudocluster (e.g. # of child points/10,000).

Model Geographies

As described in the main text, we fit five models per indicator—one per region. We used the regional definitions from the Global Burden of Disease.[10] To reduce edge effects, we included all data within one degree of the regional boundaries during the model fit stage, while the boundaries were clipped to the proper region during prediction/map making.

MBG model description, formula, terms and priors

We used a Bayesian hierarchical framework and a logit link to model our binomial count data—with separate models per region and indicator (each event category and symptom group. The model formula is provided below:

$$C_i|p_i,\ N_i \sim Binomial(p_i,\ N_i)$$

$$logit(p_i) = \beta_0 + \boldsymbol{X_i\beta} + \epsilon_{GP_i} + \epsilon_{ctry_i} + \epsilon_i$$

$$\epsilon_{ctry_i} \sim N(0, \sigma^2_{ctry})$$

$$\epsilon_i \sim N(0, \sigma^2_{ug})$$

$$\epsilon_{GP}|\Sigma_{\text{space}},\ \Sigma_{time} \sim GP(0,\ \Sigma_{\text{space}} \otimes\ \Sigma_{time})$$

$$\Sigma_{\text{space}} = \frac{2^{1-\nu}}{\tau \times \Gamma(\nu)} \times (\kappa \boldsymbol{D})^{\nu} \times \mathrm{K}_{\nu}(\kappa \boldsymbol{D})$$

$$\Sigma_{time\ j,k} = \rho^{|\ t_k - t_j|}$$

Parameter id	Parameter	Explanation
1	i	(pseudo)Cluster index
2	N_i	Sample size
3	C_i	Number of children with the modelled event category/symptom group
4	p_i	Probability/prevalence of disease
5	$\boldsymbol{X_i}$	Covariates
6	$\boldsymbol{\beta}$	Fixed effect coefficients
7	ϵ_{GP_i}	Spatiotemporally correlated errors
8	Σ_{space}	Spatial covariance modelled with an isotropic and stationary Matérn function. For the stochastic partial differential equation approximation, we set the maximum edge length for internal triangles to .2 degrees and used 5 degrees for the outer triangles in the mesh.
9	Σ_{time}	Temporal covariance, modelled with an AR1 function
10	ϵ_{ctry_i}	Country random effect
11	ϵ_i	Nugget effect—data point specific measurement error

*Parameters 1-5, 7 and 11 are also indexed spatially and temporally.

We used the following priors in all of the models:

$$\beta_0 \sim N(\mu = 0, \sigma^2 = 2^2),$$

$$\boldsymbol{\beta} \sim^{iid} N(\mu = 0, \sigma^2 = 2^2),$$

$$log\left(\frac{1+\rho}{1-\rho}\right) \sim N(\mu = 0, \sigma^2 = 1/0.15)$$

$$log\left(\frac{1}{\sigma^2_{ctry}}\right) \sim loggamma(\alpha = 1, \gamma = 2)$$

$$log\left(\frac{1}{\sigma^2_{nug}}\right) \sim loggamma(\alpha = 1, \gamma = 2).$$

$$\theta_1 = \log(\tau) \sim N(\mu_{\theta_1}, \sigma^2_{\theta_1})$$

$$\theta_2 = \log(\kappa) \sim N(\mu_2, \sigma^2_{\theta_2})$$