Role of malnutrition and parasite infections in the spatial variation in children's anaemia risk in northern Angola

Ricardo J. Soares Magalhães[1], Antonio Langa[2], João Mário Pedro[2,3], José Carlos Sousa-Figueiredo[4,5], Archie C.A. Clements[1], Susana Vaz Nery[2]

[1]*University of Queensland, Infectious Disease Epidemiology Unit, School of Population Health, Herston, Queensland, Australia;* [2]*Centro de Investigação em Saúde em Angola, Caxito, Hospital Provincial do Bengo, Caxito, Angola;* [3]*Escola Superior de Tecnologia da Saúde de Lisboa, 1990-096 Lisboa, Portugal;* [4]*Disease Control Strategy Group, Liverpool School of Tropical Medicine, Liverpool, United Kingdom;* [5]*Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom*

Abstract. Anaemia is known to have an impact on child development and mortality and is a severe public health problem in most countries in sub-Saharan Africa. We investigated the consistency between ecological and individual-level approaches to anaemia mapping by building spatial anaemia models for children aged ≤15 years using different modelling approaches. We aimed to (i) quantify the role of malnutrition, malaria, *Schistosoma haematobium* and soil-transmitted helminths (STHs) in anaemia endemicity; and (ii) develop a high resolution predictive risk map of anaemia for the municipality of Dande in northern Angola. We used parasitological survey data for children aged ≤15 years to build Bayesian geostatistical models of malaria (PfPR$_{≤15}$), *S. haematobium*, *Ascaris lumbricoides* and *Trichuris trichiura* and predict small-scale spatial variations in these infections. Malnutrition, PfPR$_{≤15}$, and *S. haematobium* infections were significantly associated with anaemia risk. An estimated 12.5%, 15.6% and 9.8% of anaemia cases could be averted by treating malnutrition, malaria and *S. haematobium*, respectively. Spatial clusters of high risk of anaemia (>86%) were identified. Using an individual-level approach to anaemia mapping at a small spatial scale, we found that anaemia in children aged ≤15 years is highly heterogeneous and that malnutrition and parasitic infections are important contributors to the spatial variation in anaemia risk. The results presented in this study can help inform the integration of the current provincial malaria control programme with ancillary micronutrient supplementation and control of neglected tropical diseases such as urogenital schistosomiasis and STH infections.

Keywords: anaemia, malnutrition, malaria, helminth infections, risk mapping, Angola.

Introduction

Childhood anaemia is a severe public health problem in sub-Saharan Africa affecting 67% or 83.5 million children in the region (WHO, 2008). Anaemia in children is associated with an increased risk of death, and may impair cognitive development (Nelson, 1996; Grantham-McGregor and Ani, 2001; Olson et al., 2009), growth (Lawless et al., 1994) and immune function (Oppenheimer, 2001). It has a multifactorial aetiology and the major contributors include non-infectious and infectious causes. Non-infectious causes

include malnutrition (Fishman et al., 2000; Semba and Bloem, 2002; Kazembe et al., 2007; Allen and Peerson, 2009), inherited haemoglobinopathies (Morris et al., 2006) and thalassemias (Wambua et al., 2006). Infectious causes include malaria (Crawley, 2004), urogenital schistosomiasis (caused by *Schistosoma haematobium*), soil-transmitted helminthiasis (STH) caused by *Ascaris lumbricoides*, *Trichuris trichiura* and hookworm (Brooker et al., 1999; Stoltzfus et al., 2000; Bates et al., 2007), human immunodeficiency virus (HIV) infection (Adias et al., 2006) and bacteraemia caused by *Steptococcus pneumoniae*, non-typhi *Salmonella* spp., *Haemophilus influenzae* type b (Means, 2000; Williams et al., 2009) and other organisms. Iron-deficiency is the most common reason for anaemia and efforts to mitigate it include the delivery of iron supplements to populations at risk (WHO, 2001). While parasitic infections such as malaria, schistosomiasis and STH contribute to iron-deficiency anaemia due to the feeding mechanisms of the parasites that lead to blood loss, the worms also cause anaemia

Corresponding author:
Ricardo J. Soares Magalhães
University of Queensland
Infectious Disease Epidemiology Unit
School of Population Health
Herston, Queensland, Australia
Tel. +61 073 365 5408; Fax +61 073 365 5442
E-mail: r.magalhaes@sph.uq.edu.au

of inflammation (Olson et al., 2009).

The prevalence of anaemia has been used as a measurable evaluation indicator of parasite control programmes because it reflects the aim of the control policy, which is to control morbidity caused by these infections (Stoltzfus et al., 1998; Guyatt et al., 2001; Brooker et al., 2004; Kabatereine et al., 2007; Koukounari et al., 2007). However, nutritional and infectious causes of anaemia are geographically variable and for efficient delivery of limited resources, the correct set of interventions needs to be targeted to population subgroups at increased risk of anaemia (Soares Magalhães and Clements, 2011a). The development of maps indicating the geographical risk profile of anaemia can help identify communities most in need and, if based on information on the major aetiologies of anaemia, allows an assessment of the risk of anaemia due to different causes (Soares Magalhães and Clements, 2011b).

In West Africa, we recently predicted the risk of anaemia in children aged ≤5 years using a spatial ecological approach to quantifying the contribution that malnutrition and parasite infections make to the overall anaemia burden using population attributable fractions (PAFs) (Soares Magalhães and Clements, 2011b). The results of this study were consistent with a previous study using individual-level data (Brooker et al., 1999). However, the estimated relative risks for these factors are prone to regression dilution bias because of the imperfect measure of infections status (a spatially predicted mean prevalence, rather than actual infection status), which may contribute to more conservative PAF estimates (Hutcheon et al., 2010). Therefore it is important to evaluate the consistency between the ecological approach described previously and a complementary individual-level approach.

Recently, a sub-national parasitological survey was conducted in Angola (Dande municipality, Bengo province) which included the measurement of haemoglobin (Hb) concentration, anthropometric indicators, malaria, urogenital schistosomiasis and STH infections (Sousa-Figueiredo et al., 2012). Using these data, we aimed to quantify the role of factors associated with the geographical variation of anaemia in children aged ≤15 years, determine the geographical distribution of anaemia using an individual-level modelling approach and provide the first high-resolution, local-scale anaemia risk map that can facilitate the planning and implementation of integrated interventions that include micronutrient supplementation and fortified foods *versus* deworming and malaria control within the region.

Materials and methods

Ethics statement

The study protocol was approved by the Angolan Ministry of Health Ethics Committee. One day before the survey, a field worker visited each household to explain the study and obtain individual written informed consent (or fingerprint for illiterates) from the selected caregiver. All households that agreed to participate were enrolled in the study after handing in the signed consent form. Participants found to be infected by *Plasmodium falciparum* were treated with artemisinin-based combination therapy; those with *S. haematobium* or *S. mansoni* infection were treated with a standard dose of praziquantel (40 mg/kg) and those found to have any STH received 400 mg albendazole (children aged <2 years were treated with 200 mg of albendazole).

Data sources

We used parasitological data from a baseline, community-based, cross-sectional survey conducted between May and August 2010 in three of five communes (Caxito, Mabubas and Úcua) in the Dande municipality, Bengo province, Angola, which are part of a demographic surveillance system (DSS) maintained by the Centre for Health Research in Angola (abbreviated CISA in Portuguese) (Costa et al., 2012; Sousa-Figueiredo et al., 2012). Detailed information about the study area, sampling design and survey management are given in the paper by Sousa-Figueiredo et al. (2012). Briefly, a total of 972 households, distributed between 36 hamlets, were selected. Study participants included 960 mothers or caregivers (mean age 33.3 years, range 16-80 years) and 2,379 children (mean age 5.9 years, range 6 months to 15 years). Identification of malaria parasites in the blood was done microscopically by two independent technicians (double-blind), who screened blood films stained with 10% Giemsa in the field (Fleischer, 2004). Each child was measured for height to the nearest 0.1 cm and for weight to the nearest 0.5 kg and z-scores for weight-for-age (an indicator of underweight); height-for-age (an indicator of stunting); and weight-for-height (an indicator of wasting) were estimated using the 2006 World Health Organization (WHO) database for child growth standards. Finger-prick blood was collected from each participant and anaemia was diagnosed using a HemoCue photometer (HemoCue 201+ system; HemoCue, Ängelholm, Sweden) based

on the WHO Hb concentration thresholds (WHO, 2001). A single stool and a single urine sample were collected from each participant. Eggs of *A. lumbricoides* and *T. trichiura* were investigated using two Kato-Katz thick smears (Katz et al., 1972) and urogenital schistosomiasis was diagnosed as done by Lwambo et al. (1997) by detecting micro-haematuria in urine using urine reagent strips (Hemastix; Bayer, Leverkusen, Germany). Information relating to the mother or caregiver and their children was collected through a questionnaire, which included age, sex, literacy, occupation, access to healthcare, history of previous anthelminthic treatment as well as questions related to malaria such as bed-net ownership and utilization. Household information regarding type of flooring, ceiling and water, sanitation and hygiene (WASH) indicators was obtained from the CISA-DSS data warehouse.

Procedures for geolocation of households and details of the data related to the physical environment used in the study are described in the paper by Costa et al. (2012). In brief, we used the geographical centre of the sector (the lower subdivision of the hamlet) as the geographical unit for the analysis. This enabled 95% of individuals included in the survey to be successfully georeferenced to 43 locations. Electronic data for land surface temperature and rainfall for a 1×1 km grid cell resolution were obtained from the WorldClim data warehouse and the distance to lakes, rivers and irrigation canals was extracted for each sector in a geographical information system (GIS) using ArcView version 10.0 (ESRI, Redlands, USA). The geographical centre of the sector was linked to the environmental data to obtain values of the environmental variables using the spatial overlay procedure.

Data analysis framework

We built Bernoulli logistic regression, multivariable models for each parasite infection using the environmental variables as covariates and investigated residual, spatial variation using semivariograms (Data S1). This assessment revealed considerable residual spatial clustering, which justified modelling of second-order, spatial variation using model-based geostatistics (MBG) (Fig. S1). The subsequent analyses were conducted in two phases (Data S1; Fig. S2). In phase 1, we investigated the geographical variation of parasite infections in children aged ≤15 years using MBG (Diggle et al., 1998). In phase 2, we evaluated the impact of imprecise measurement of parasite infections on their effect sizes in relation to anaemia; esti-

mated the population attributable fraction (PAF) of anaemia due to malnutrition and parasite infections; and predicted the geographical variation of anaemia in children aged ≤15 years in the region using MBG.

Spatial models

In phase 1, the models of malaria, urogenital schistosomiasis, ascariasis and trichuriasis included the same sets of environmental covariates and a geostatistical random effect (Data S1). The prediction of malaria for the study area was made for children aged ≤15 years and therefore defined as the *P. falciparum* parasite rate for his age group ($Pf\,PR_{\leq15}$). The geostatistical random effect has two parameters: the tendency for clustering (as assessed by the variance of spatial random effect) and the cluster radii (as assessed by the parameter Phi (φ) which indicates the rate of decay of spatial autocorrelation). Note that, φ, is measured in decimal degrees and $3/\varphi$ determines the size of the clusters in km (one decimal degree is approximately 111 km at the equator). In phase 2, the Bayesian geostatistical models of anaemia prevalence included age, sex, nutritional and infectious contributors of anaemia, and a geostatistical random effects. Covariate effects were summarised using the mean and 95% credible intervals (representing the range of values that contains the true value with a probability of 95%). The geostatistical random effects modelled spatial correlation as a function of the separating distance between locations.

In phase 2, we built two anaemia models. In model 1 (the individual-level approach), we used the observed parasite infection status as the measure of exposure; in model 2 (the ecological approach), the values of the predicted parasite infections at sampled locations were added as random variables (using their mean and standard deviation (SD) from the phase 1 model) in a MBG model of anaemia, instead of a fixed mean (Soares Magalhães, 2011). This method incorporates the geographical uncertainty in predicted prevalence rates of parasite infections. The impact of incorporating this uncertainty was shown by comparing the overall SD in infection risk across the study area to the overall SD from the simpler model using mean parasite infections.

Spatial prediction of anaemia was performed using model 2 on a 0.005×0.005 decimal-degree grid (approximately 500×500 m) covering the municipality of Dande (across a total area of 5,500 km²). Model 2 was selected for spatial prediction because the predictions of parasite prevalence could be used to inter-

polate anaemia risk across the study area. An anaemia risk map for stunted boys (the subgroup with the highest risk of anaemia) for the entire study area was constructed using ArcGIS version 10.0. All models were developed in WinBUGS 1.4 (Medical Research Council, Cambridge, UK and Imperial College London, London, UK).

Model validation

To determine the discriminatory performance of model predictions relative to observed prevalence thresholds the area under the curve (AUC) of the receiver operating characteristic was used. To assess the predictive performance of parasite prevalence models, we used a prevalence threshold of 10% (the lower prevalence threshold for mass drug administration for schistosomiasis) and in case of the anaemia model, we used a prevalence threshold of 50% (prevalence threshold defining anaemia as a public health problem). An AUC value of 0.7 was taken to indicate acceptable predictive performance (Brooker et al., 2002).

PAF of anaemia due to malnutrition and parasite infections

Estimates of the PAF for specific predictors are used to guide policymakers in planning public health interventions (Rowe et al., 2004; Rothman et al., 2008). PAF estimates in this study represent the fraction of total anaemia risk in the population caused by each aetiological factor, or alternatively, would not have occurred if the aetiological factor were eliminated, assuming the effects of other contributors (e.g. environmental variables) remain unchanged (Bruzzi et al., 1985).

We used model 1 to estimate the risk of anaemia in children aged ≤15 years attributable to malnutrition and parasite infections, because we expected that effects estimated using this approach would be less prone to regression dilution bias. Estimation proce-

dures for PAF of anaemia for malnutrition and parasite infections in children aged ≤15 years are presented in the supplementary Data S1.

Results

Dataset for analysis

We included all children with complete data on location (i.e. hamlet coordinates), demography (i.e. age and sex), Hb concentration, nutritional status (i.e. height-for-age z-score and weight-for-age z-score), and infection status (i.e. malaria, *S. haematobium*, *A. lumbricoides* and *T. trichiura*). A total of 1,884 children aged ≤15 years, comprising 955 girls and 929 boys (636 children in Caxito), 712 children in Mabubas and 536 children in Úcua) were included in the analysis (Tables 1 and 2). In the dataset for analysis, the mean age in years was 7 (SD: 3.7), the mean Hb concentration was 11.2 g/dl (range: 3.0-15.9 g/dl), and the overall prevalence of anaemia was 39.6%. The prevalence of stunting was 30.0% and that of underweight was 22.4%. The proportion of malaria was 19.1%. The prevalence of infection with *S. haematobium* was 13.0%, *A. lumbricoides* 16.4%, *T. trichiura* 10.8%, and *A. lumbricoides/T. trichiura* co-infection 3.2%.

Geographical variation in parasite infections

Environmental variables significantly associated with the parasite infections are presented in Table 3. Model results indicate that rainfall was positively associated with malaria and *A. lumbricoides* and negatively associated with *S. haematobium* and *T. trichiura*. The effect of rainfall was only significant for *S. haematobium*. At the 5% level, land surface temperature was not significantly associated with any of the parasite infections studied. Distance to irrigation canals was significantly and negatively associated with *S. haematobium* infection only. Distance to health care centres was significantly and positively associated with

Table 1. Characteristics of children aged ≤15 years by commune in Dande municipality, Angola.

District	No. of children examined (male : female)	Mean age ±SD	Mean Hb concentration (g/dl) (min - max)	Observed anaemia prevalence (%)
Caxito	309 : 327	7 ± 3.6	11.4 (3.0-15.6)	31.6
Mabubas	346 : 366	7 ± 3.8	11.3 (4.5-15.1)	39.0
Úcua	274 : 262	6 ± 3.6	10.8 (4.3-15.9)	48.9
Total	929 : 955	7 ± 3.7	11.2 (3.0-15.9)	39.6

Table 2. Characteristics of children aged ≤15 years selected for analysis by the level of anaemia (Hb <11.0 g/dl).

| Variable | Anaemia | | Total |
	No	Yes	
Hb concentration (g/dl)			
Mean	12.2	9.7	
Range (minimum-maximum)	11.0-15.9	3.0-10.9	
Age in years			
≤5	867	358	1,225
5-15	314	437	751
Sex			
Male	558	423	981
Female	623	372	995
Height-for-age z-score <-2			
No	855	467	1,322
Yes	283	279	562
Weigth-for-age z-score<-2			
No	650	536	1,186
Yes	161	181	342
Malaria			
No	1,009	588	1,597
Yes	172	207	379
S. haematobium infection			
No	1,032	688	1,720
Yes	149	107	256
A. lumbricoides infection			
No	966	686	1,652
Yes	215	109	324
T. trichiura infection			
No	1,030	732	1,762
Yes	151	63	214

malaria and distance to rivers was significantly and negatively associated with malaria. The tendency for clustering was highest for *A. lumbricoides* and lowest for *T. trichiura*. After accounting for the effect of covariates, the radii of the clusters of malaria, *S. haematobium* infection, *A. lumbricoides* and *T. trichiura* were approximately 24 km, 22 km, 26 km and 27 km, respectively.

We found large clusters of high risk of malaria infection (prevalence >50%) in a region covering the North and most of the eastern areas of the Dande municipality (Fig. S3 A). More circumscribed areas of high risk of malaria were found between the communes of Quicabo and Mabubas and between the communes of Mabubas and Úcua in central Dande. There is also a large elongated cluster predicted in the southern part of the province in close proximity to inland water bodies in those areas. With respect to *S. haematobium* infection, the predicted prevalence reached the highest level (>50%) in two large clusters in eastern Barra do Dande and in southern Úcua in close proximity with the inland water bodies (Fig. S3 B). The predicted prevalence of *T. trichiura* (Fig. S3 C) and *A. lumbricoides* (Fig. S3 D) was low (<10%) across most of the study area. The final predictive models showed acceptable predictive capability (i.e. AUC >0.7) (Table 3).

Spatial models of anaemia

Age in years was significantly and negatively associated with anaemia risk but sex was not significantly associated with anaemia risk or stunting in both

Table 3. Spatial effects of the physical environment on parasite infection (mean values), in children aged ≤15 years in Dande municipality, Angola.

Variable	Malaria (95% CrI*)	*S. haematobium* (95% CrI*)	*A. lumbricoides* (95% CrI*)	*T. trichiura* (95% CrI*)
Rainfall**	0.45 (-0.50, 1.57)	-0.89(-1.97, -0.07)	0.32 (-1.25, 1.84)	-0.05 (-0.59, 0.50)
Temperature**	-0.29 (-1.18, 0.52)	0.47 (-0.35, 1.33)	0.14 (-0.92, 1.30)	-0.12 (-0.64, 0.43)
Distance to irrigation canals**	0.78 (-0.60, 2.05)	-1.43 (-2.56, -0.36)	0.60 (-1.05, 1.99)	0.04 (-0.54, 0.62)
Distance to health care centres**	0.78 (0.06, 1.78)	-0.005 (-0.85, 0.99)	0.12 (-1.25, 1.36)	0.08 (-0.45, 0.58)
Distance to rivers**	-0.95 (-1.90, -0.03)	0.11 (-0.82, 1.04)	0.20 (-1.02, 1.55)	-0.18 (-0.76, 0.39)
Intercept	-1.84 (-3.30, 0.70)	-1.66 (-2.84, -0.32)	-2.66 (-4.37, -0.29)	-2.11 (-2.81, -1.34)
Spatial autocorrelation decay rate (φ)	12.68 (2.97, 19.70)	13.66 (4.83, 19.71)	12.96 (2.95, 19.74)	12.55 (3.824, 19.52)
Variance of spatial random effect	4.61 (1.25, 14.72)	3.11 (1.25, 7.62)	5.89 (1.95, 18.35)	0.65 (0.11, 2.06)
Area under the ROC	0.79 (0.73, 0.85)	0.81 (0.75, 0.89)	0.78 (0.72, 0.84)	0.76 (0.71, 0.81)

*CrI, credible interval;
** Variables were standardised to have mean = 0; and standard deviation = 1

Table 4. Spatial effects anaemia in children aged ≤15 years in Dande municipality, Angola.

Variable	Model 1 Posterior mean (95% CrI*)		Model 2 Posterior mean (95% CrI*)	
Age	-0.22	(-0.26, -0.19)	-0.22	(-0.25, -0.18)
Female (*versus* male)	-0.17	(-0.38, 0.04)	-0.17	(-0.38, 0.03)
Height-for-age Z-score <-2	0.38	(0.15, 0.61)	0.37	(0.13, 0.60)
Prevalence of PfPR ≤15	0.68	(0.38, 0.99)	0.93	(-1.72, 3.82)
Prevalence of *S. haematobium* infection	0.60	(0.26, 0.94)	1.16	(-1.32, 3.40)
Prevalence of *A. lumbricoides* infection	-0.16	(-0.45, 0.13)	1.40	(-2.24, 4.99)
Prevalence of *T. trichiura* infection	-0.27	(-0.61, 0.06)	0.02	(-8.41, 7.98)
Intercept	0.95	(0.40, 1.61)	0.30	(-1.19, 3.21)
Rate of decay of spatial autocorrelation (φ)	14.7	(5.36, 19.84)	14.49	(5.36, 19.78)
Variance of spatial random effect	0.41	(0.06, 1.26)	0.76	(0.20, 1.98)

*CrI, credible interval

anaemia models tested (Table 4). While *Pf* PR$_{\leq 15}$, and *S. haematobium* infections were significantly and positively associated with anaemia risk in model 1, none of these variables was significantly associated with anaemia in model 2. Neither the prevalence of *A. lumbricoides* nor *T. trichiura* infection was significantly associated with risk of anaemia in any of the models. Using model 2, we were able to predict prevalence of anaemia being greater than 80% with an AUC of 0.82 (95% CI: 0.73-0.89). Using this model we found that the risk of anaemia was consistently high across the entire study area, with maximal prevalence (>85%) in small foci in Caxito, central Mabubas and in an elongated cluster in south Úcua in association with inland water bodies (Fig. 1). Larger foci of prevalence of anaemia between 80-85% were also predicted for northern Quicabo and areas adjacent to the inland water bodies in eastern Barra do Dande. The average size of anaemia clusters (given by φ) was approximately 22 km. Incorporating measurement error of the exposure variables (i.e. including predicted parasite prevalence as random variables as opposed to just considering the mean prediction) increased the mean SD across all unsampled locations by 8.7% in the anaemia model.

Risk of anaemia attributable to malnutrition and parasite infections

Using model 1, our results indicated that the estimated risk of anaemia in children aged ≤15 years

attributable to stunting was 12.4%, *P. falciparum* was 15.6% and *S. haematobium* was 9.7%.

Discussion

Previous work in West Africa investigated the role of spatial variation in malaria and helminth infections in pre-school anaemia risk based on an ecological approach to anaemia modelling (Soares Magalhães and Clements, 2011b). In the current study we adopted an individual-level approach to anaemia modelling whereby children's infection status for malaria, urogenital schistosomiasis and STH, respectively, was used as a measure of exposure. We also built a model in which infection status was assigned with measurement error (by spatial overlaying the anaemia survey locations to the predicted parasite surfaces) and compared the results with the individual-level model. Our results confirm the hypothesis that measurement error in an exposure variable is likely to bias the estimates of regression coefficients towards the null (Hutcheon et al., 2010). In this study, none of the effects of parasite infections in the model with measurement error were statistically significant. This also suggests that the results we obtained in our study in West Africa using an ecological approach are likely to underestimate the role of malnutrition, malaria and helminth infections in children aged <5 years. Importantly, the resulting predictive anaemia map with measurement error reported in this study (Fig. 1) is capable of identifying clusters of high anaemia risk suggesting that in
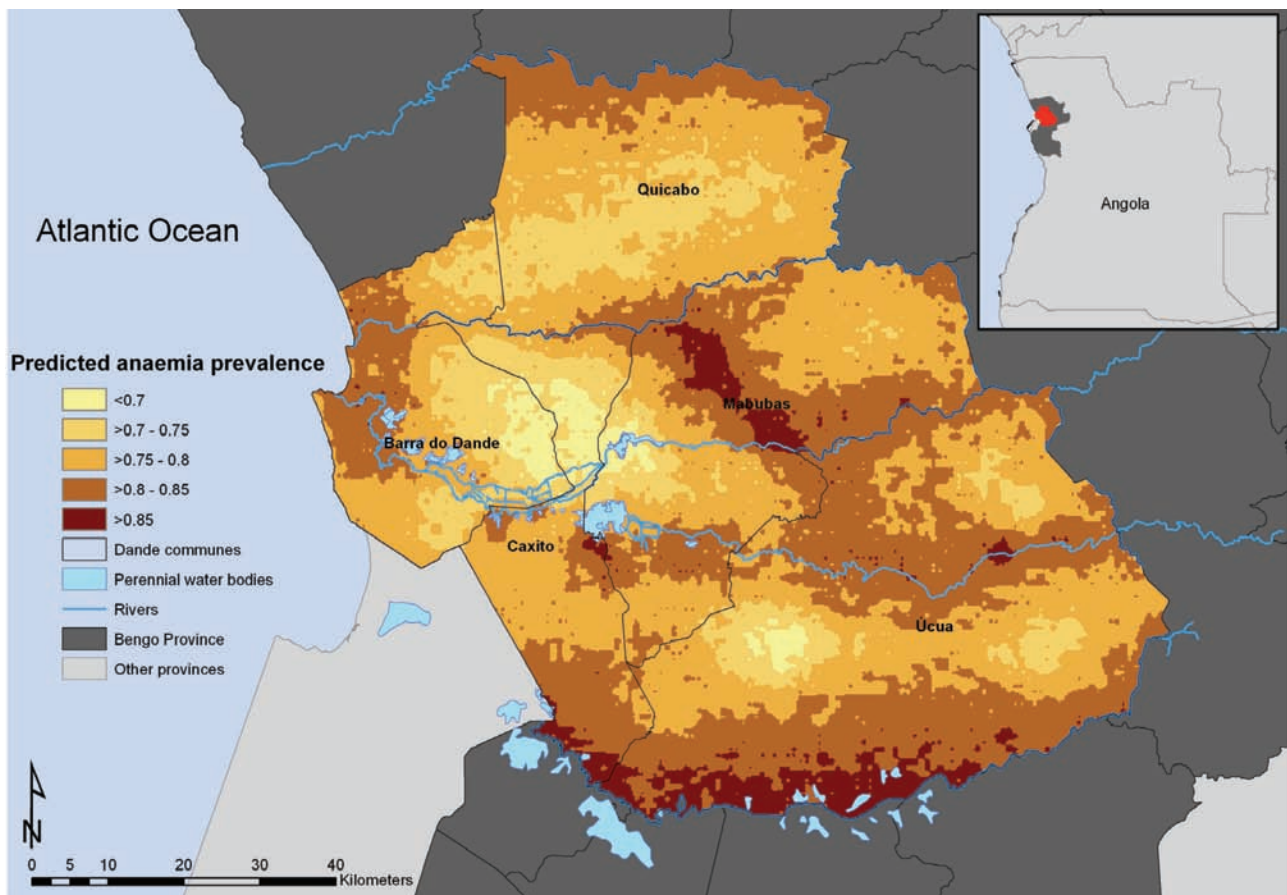
Fig. 1. Predicted spatial distribution of anaemia (Hb <11.0 g/dl) for younger boys in Dande municipality, Angola.

the absence of individually collected data, an ecological approach may be a valid approach to identify communities at the highest risk for anaemia.

Extending anaemia modelling to include school-aged children is important as current guidelines for the prevention and control of human helminthiasis target this age group (WHO, 2002). In addition, the epidemiology of anaemia is likely to be different between pre-school and school-aged children as the age prevalence profiles for parasite infections such as schistosomiasis and STH indicate a greater risk in school-aged children (Drake and Bundy, 2001; Holland and Kennedy, 2002; Hotez et al., 2006). Thus, including a wider age group of children provided a more complete assessment of the effect of nutritional and infections contributors to anaemia in children.

Using a spatial approach, we confirm the previous finding that malaria plays a central role in anaemia burden in this region of Angola (Sousa-Figueiredo et al., 2012). Our results show that almost 16% of anaemia cases in children aged ≤15 years could have been averted in 2010 by eliminating malaria in the population. These results are supported by earlier findings in West Africa using an ecological-approach (PAF of 15% among infected pre-school-aged children) (Soares Magalhães and Clements, 2011b) and in Kenya using individual-level data (PAF of 14% among infected pre-school-aged children and 7% in the whole population) (Brooker et al., 1999). The same model also underlines the role of malnutrition in the anaemia burden in that the proportion of anaemia attributable to malnutrition was 12%. Surprisingly, this figure is lower than that reported by our previous study in West Africa using an ecological-approach for children aged ≤5 years where the PAF of anaemia due to malnutrition was estimated to be 35% (Soares Magalhães and Clements, 2011b). Despite the mean prevalence of malnutrition in the population being almost two-fold higher than that of malaria, the small attributable risk for malnutrition is because its relative risk was modest compared to that of malaria.

The PAF of anaemia for *S. haematobium* infection (9.8%) was higher than that reported in the West Africa study indicating that *S. haematobium* infection

plays an important role in anaemia risk in the study area. The higher effect for *S. haematobium* may be due to the fact that in this study, in addition to children aged <5 years, we included children aged 5-15 years, who are known to be at increased risk of *S. haematobium* compared to children aged <5 years (the age group included in the West Africa study).

This study provides important new information about sub-national priority areas for targeting anaemia control in the region. The resulting predictive map of anaemia is based on an ecological approach in that mapped outputs from spatial models of malaria, *S. haematobium* and STHs were used. While in the spatial model of urogenital schistosomiasis the effects of the physical environment in the model are consistent with the known epidemiology of *S. haematobium* infection (Brooker et al., 2009), none of the environmental variables were statistically significant at the 5% level in the STH models, which is surprising. This could reflect the relatively homogeneous environment of this small study area, or a dominance of the effects of improper WASH over the physical environment at small spatial scales (Mara and Feachem, 1999). In order to generate the predictive anaemia map, we included the prediction uncertainty around the predicted mean of parasite infection covariates by modelling the distribution of probable values (Soares Magalhães et al., 2011). Taking explicit consideration of the uncertainty of mapped values of parasite infection surfaces has contributed to a more detailed handling of prediction uncertainty.

Our predictive map of anaemia indicates that anaemia control should be prioritised to inland rural communities within the high endemicity areas identified in our maps (prevalence >85%). These areas correlated with larger areas where malaria risk is at its highest (prevalence >50%). Furthermore, high risk areas of *S. haematobium* (>50%) overlapped with high risk areas of malaria (>50%), particularly between the Caxito and Mabubas communes and southern Úcua. These areas present opportunities for integrating malaria interventions with praziquantel delivery to reduce anaemia prevalence. It would be useful to include distribution of micronutrients for the treatment of severe anaemia cases as part of deworming programmes, particularly in areas with high prevalence rates of helminth infections. In addition, when financially possible, WASH should be provided with the aim of reducing the burden of anaemia attributable to helminth infections (Stoltzfus et al., 1997; WHO, 2006). Taken together, our results suggest that resources for the treatment of anaemia, such as iron

supplementation, malaria control, deworming and blood for emergency transfusion, should be prioritised towards populations located in the clusters of high anaemia risk identified in this study.

We acknowledge some inherent limitations in our study. In PAF estimation the multivariable model needs to be as complete as possible; if one or several factors act as confounders in the association between exposure and disease, the crude PAF estimates can be biased and there is a need for adjustment when estimating the PAF (Benichou, 1991). Regression models allow taking into account adjustment factors as well as interaction of exposure with some or all adjustment factors (Benichou, 2001). We have attempted statistical control of confounding by adjusting our analysis for age and sex. However, even if one uses adjusted estimates of the relative risk, PAF estimates can be biased in the presence of unaccounted confounding factors and overestimation of PAF can occur (Bruzzi et al., 1985; Ruckinger et al., 2009). Haemoglobinopathies and thalassemias, which were not incorporated in the models, are important inherited haematological conditions, particularly in populations in West Africa (Weatherall and Clegg, 2001) but predictive surfaces for the sickle cell trait have only recently become available (Piel et al., 2010).

Conclusions

Our approach using individual-level data collected from children aged ≤15 years revealed significant spatial variation in anaemia risk explained by malnutrition and different infections. While an ecological approach to anaemia modelling is likely to underestimate the role of different contributors our results show that, in the absence of individually collected data, it may be a valid approach to identify communities at highest risk of anaemia. While the results presented in this study can facilitate the integration of the current provincial malaria control programme with ancillary micronutrient supplementation and control of urogenital schistosomiasis and STH infection, further research is needed to quantify the most cost-effective delivery platform for the effective reduction of childhood anaemia.

Acknowledgements

References

Adias TC, Uko E, Erhabor O, 2006. Anaemia in human immun-odeficiency virus infection: a review. Niger J Med 15, 203-206.

Allen LH, Peerson JM, 2009. Impact of multiple micronutrient versus iron-folic acid supplements on maternal anemia and micronutrient status in pregnancy. Food Nutr Bull 30, 527-532.

Bates I, McKew S, Sarkinfada F, 2007. Anaemia: a useful indicator of neglected disease burden and control. PLoS Med 4, e231.

Benichou J, 1991. Methods of adjustment for estimating the attributable risk in case-control studies: a review. Stat Med 10, 1753-1773.

Benichou J, 2001. A review of adjusted estimators of attributable risk. Stat Methods Med Res 10, 195-216.

Brooker S, Hay SI, Bundy DAP, 2002. Tools from ecology: useful for evaluating infection risk models? Trends Parasitol 18, 70-74.

Brooker S, Kabatereine NB, Gyapong JO, Stothard JR, Utzinger J, 2009. Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. Parasitology 136, 1707-1718.

Brooker S, Peshu N, Warn PA, Mosobo M, Guyatt HL, Marsh K, Snow RW, 1999. The epidemiology of hookworm infection and its contribution to anaemia among pre-school children on the Kenyan coast. Trans R Soc Trop Med Hyg 93, 240-246.

Brooker S, Whawell S, Kabatereine NB, Fenwick A, Anderson RM, 2004. Evaluating the epidemiological impact of national control programmes for helminths. Trends Parasitol 20, 537-545.

Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C, 1985. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 122, 904-914.

Costa MJ, Rosário E, Langa A, António G, Bendriss A, Vaz Nery S, 2012 Setting up a demographic surveillance system in Northern Angola. African Pop Stud 26, 2.

Crawley J, 2004. Reducing the burden of anemia in infants and young children in malaria-endemic countries of Africa: from evidence to action. Am J Trop Med Hyg 71 (suppl. 2), 25-34.

Diggle P, Moyeed R, Tawn J, 1998 Model-based Geostatistics (with discussion). Applied Statistics. 48, 299-350.

Drake LJ, Bundy DAP, 2001. Multiple helminth infections in children: impact and control. Parasitology 122, S73-81.

Fishman SM, Christian P, West KP, 2000. The role of vitamins in the prevention and control of anaemia. Public Health Nutr 3, 125-150.

Fleischer B, 2004. 100 years ago: Giemsa's solution for staining of plasmodia. Trop Med Int Health 9, 755-756.

Grantham-McGregor S, Ani C, 2001. A review of studies on the effect of iron deficiency on cognitive development in children. J Nutr 131, 649-666.

Guyatt HL, Brooker S, Kihamia CM, Hall A, Bundy DAP, 2001. Evaluation of efficacy of school-based anthelmintic treatments against anaemia in children in the United Republic of Tanzania. Bull World Health Organ 79, 695-703.

Holland CV, Kennedy MW, 2002. The geohelminths: *Ascaris, Trichuris* and hookworm. Kluwer Academic Publishers.

Hotez PJ, Bundy DAP, Beegle K, Brooker S, Drake L, de Silva N, Montresor A, Engels D, Jukes M, Chitsulo L, Chow J, Laxminarayan R, Michaud CM, Bethony J, Correa–Oliveira R, Xiao SH, Fenwick A, Savioli L, 2006. Helminth infections: soil–transmitted helminth infections and schistosomiasis disease control priorities in developing countries. New York: Oxford University Press.

Hutcheon JA, Chiolero A, Hanley JA, 2010. Random measurement error and regression dilution bias. BMJ 340, c2289.

Kabatereine NB, Brooker S, Koukounari A, Kazibwe F, Tukahebwa EM, Fleming FM, Zhang Y, Webster JP, Stothard JR, Fenwick A, 2007. Impact of a national helminth control programme on infection and morbidity in Ugandan schoolchildren. Bull World Health Organ 85, 91-99.

Katz N, Chaves A, Pellegrino J, 1972. A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. Rev Inst Med Trop São Paulo 14, 397-400.

Kazembe LN, Appleton CC, Kleinschmidt I, 2007. Geographical disparities in core population coverage indicators for roll back malaria in Malawi. Int J Equity Health 6, 5.

Koukounari A, Fenwick A, Whawell S, Kabatereine NB, Kazibwe F, Tukahebwa EM, Stothard JR, Donnelly CA, Webster JP, 2006. Morbidity indicators of *Schistosoma mansoni*: relationship between infection and anemia in Ugandan schoolchildren before and after praziquantel and albendazole chemotherapy. Am J Trop Med Hyg 75, 278-286.

Lawless JW, Latham MC, Stephenson LS, Kinoti SN, Pertet AM, 1994. Iron supplementation improves appetite and growth in anemic Kenyan primary school children. J Nutr 124, 645-654.

Lwambo NJ, Savioli L, Kisumku UM, Alawi KS, Bundy DA, 1997. Control of *Schistosoma haematobium* morbidity on Pemba Island: validity and efficiency of indirect screening tests. Bull World Health Organ 75, 247-252.

Mara DD, Feachem RGA, 1999. Water- and excreta-related diseases: unitary environmental classification. J Environ Eng 125, 334-339.

Means RT, 2000. The anaemia of infection. Best practice and

research. Clin Haematol 13, 151-162.

Morris CR, Singer ST, Walters MC, 2006. Clinical hemoglo-binopathies: iron, lungs and new blood. Curr Opin Hematol 13, 407-418.

Nelson M, 1996. Anaemia in adolescent girls: effects on cognitive function and activity. Proc Nutr Soc 55, 359-367.

Olson CL, Acosta LP, Hochberg NS, Olveda RM, Jiz M, McGarvey ST, Kurtis JD, Bellinger DC, Friedman JF, 2009. Anemia of inflammation is related to cognitive impairment among children in Leyte the Philippines. PLoS Negl Trop Dis 3, e533.

Oppenheimer SJ, 2001. Iron and its relation to immunity and infectious disease. J Nutr 131, 616-633.

Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Williams TN, Weatherall DJ, Hay SI, 2010. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. Nat Commun 1, 104.

Rothman KJ, Greenland S, Lash TL, 2008. Modern epidemiology: Lippincott, Williams and Wilkins.

Rowe AK, Powell KE, Flanders WD, 2004. Why population attributable fractions can sum to more than one. Am J Prev Med 26, 243-249.

Ruckinger S, von Kries R, Toschke AM, 2009. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. BMC Med Res Methodol 9, 7.

Semba RD, Bloem MW, 2002. The anemia of vitamin A deficiency: epidemiology and pathogenesis. Eur J Clin Nutr 56, 271-281.

Soares Magalhaes RJ, Clements ACA, 2011a. Spatial heterogeneity of haemoglobin concentration in preschool-age children in sub-Saharan Africa. Bull World Health Organ 89, 459-468.

Soares Magalhães RJ, Clements ACA, 2011b. Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition malaria and helminth infections in West Africa. PLoS Med 8, e1000438.

Soares Magalhães RJ, Barnett AG, Clements ACA, 2011. Geographical analysis of the role of water supply and sanitation in the risk of helminth infections of children in West Africa. Proc Natl Acad Sci USA 108, 20084-20089.

Sousa-Figueiredo JC, Gamboa D, Pedro JM, Fançony C, Langa AJ, Soares Magalhães RJ, Stothard JR, Nery SV, 2012. Epidemiology of malaria, schistosomiasis, geohelminths, anemia and malnutrition in the context of a demographic surveillance system in northern Angola. PLoS One 7, e33189.

Stoltzfus RJ, Albonico M, Chwaya HM, Tielsch JM, Schulze KJ, Savioli L, 1998. Effects of the Zanzibar school-based deworming program on iron status of children. Am J Clin Nutr 68, 179-186.

Stoltzfus RJ, Chwaya HM, Montresor A, Albonico M, Savioli L, Tielsch JM, 2000. Malaria, hookworms and recent fever are related to anemia and iron status indicators in 0- to 5-y old Zanzibari children and these relationships change with age. J Nutr 130, 1724-1733.

Stoltzfus RJ, Dreyfuss ML, Chwaya HM, Albonico M, 1997. Hookworm control as a strategy to prevent iron deficiency. Nutr Rev 55, 223-232.

Wambua S, Mwangi TW, Kortok M, Uyoga SM, Macharia AW, Mwacharo JK, Weatherall DJ, Snow RW, Marsh K, Williams TN, 2006. The effect of alpha+-thalassaemia on the incidence of malaria and other diseases in children living on the coast of Kenya. PLoS Med 3, e158.

Weatherall DJ, Clegg JB, 2001. Inherited haemoglobin disorders: an increasing global health problem. Bull World Health Organ 79, 704-712.

WHO, 2001. Iron deficiency anaemia: assessment, prevention and control. Geneva: World Health Organization (WHO/NHD/01.3). WHO Technical Report Series 912.

WHO, 2002. Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of a WHO expert committe. WHO Tech Rep Ser 912, 1-57.

WHO, 2006. Iron supplementation of young children in regions where malaria tansmission is intense and infectious disease highly prevalent. Geneva: World Health Organization.

WHO, 2008. Worldwide prevalence of anaemia 1993-2005. In: WHO Global Database on Anaemia (Benoist, McLean, Egli, Cogswell, eds). Geneva: World Health Organization.

Williams TN, Uyoga S, Macharia A, Ndila C, McAuley CF, Opi DH, Mwarumba S, Makani J, Komba A, Ndiritu MN, Sharif SK, Marsh K, Berkley JA, Scott JA, 2009. Bacteraemia in Kenyan children with sickle-cell anaemia: a retrospective cohort and case-control study. Lancet 374, 1364-1370.

Supplementary Data

Residual spatial dependence in parasite infections

Fixed-effects Bernoulli regression models of parasite prevalence were developed in a frequentist statistical software package (Stata version 10.1, Stata Corporation, College Station, TX) to assess the residual spatial variation unaccounted by the variables of the physical environment. The assessment was conducted by extracting the residuals of each parasite model and investigating spatial dependence using semivariograms, developed using the package geoR of the statistical software R. The residual semivariograms for each parasite are shown in Supplementary Fig. S1.

Data analysis framework

The analysis was conducted in two phases (Supplementary Fig. S2): In Phase 1, we investigated the geographical variation of parasite infections in children aged ≤15 years. This involved the development of Bayesian geostatistical models of malaria, urogenital schistosomiasis (as indicated by haematuria), ascariasis and trichuriasis and the generation of predictive maps of each parasite infection using model-based geostatistics. In Phase 2, we aimed to estimate the PAF of anaemia due to malnutrition and parasite infections and to predict the geographical variation of anaemia in children aged ≤15 years in the region. With that in mind, first we evaluated the impact of imprecise measurement of parasite infections on their effect
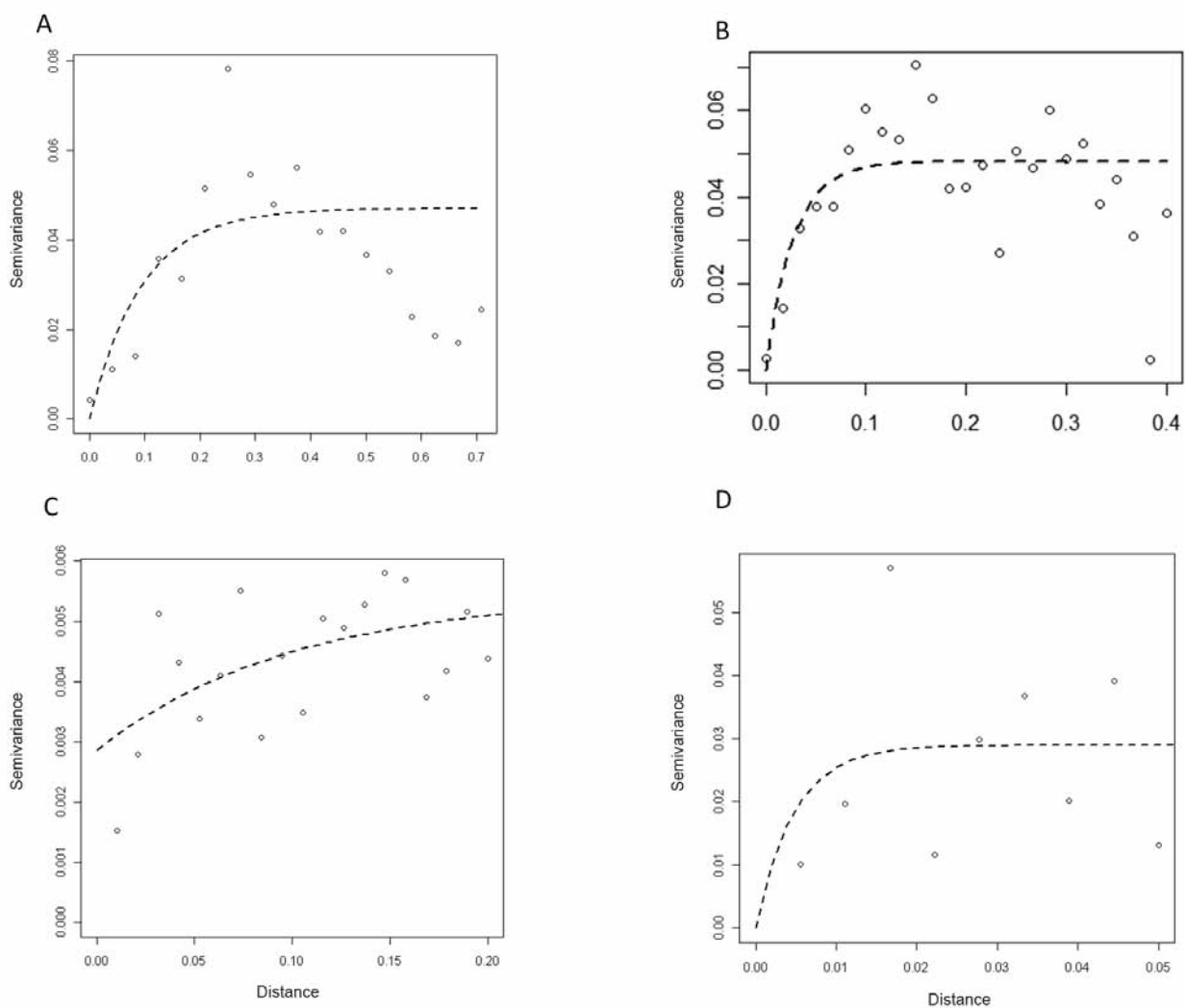

Fig. S1. Residual semivariograms for non-spatial prevalence models of malaria (A), *S. haematobium* (B), *T. trichiura* (C) and *A. lumbricoides* (D).
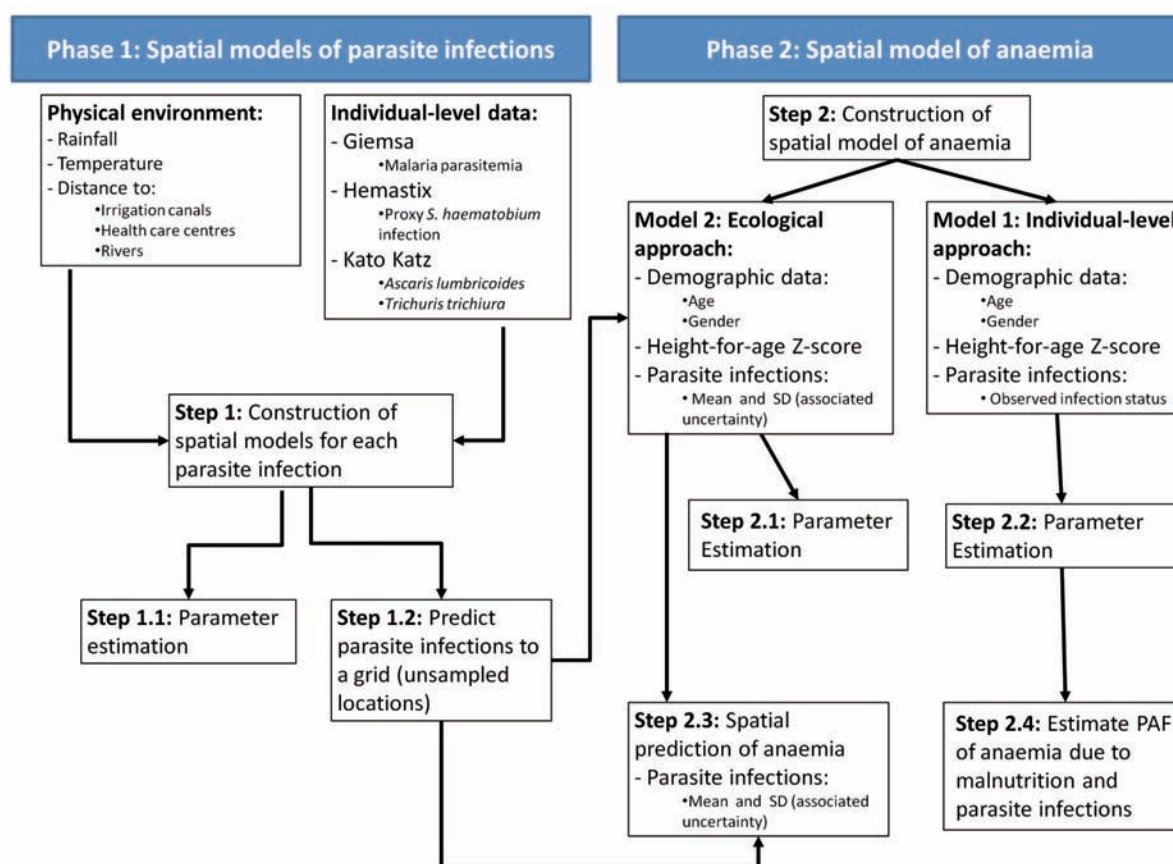
Fig. S2. Flow diagram showing the data sources and analytical steps of the analysis.

sizes in relation to anaemia by comparing two modeling approaches. In Model 1 we used an individual-level approach whereby we used the observed parasite infection status obtained in the survey as a measure of exposure; in Model 2 we used an ecological approach whereby the mean and standard deviation of the predicted prevalence of each parasite infection estimated in Phase1 were extracted to survey locations and used as measures of exposure. Secondly, we estimated PAF of anaemia due to malnutrition and parasite infections based on the results of Model 1. Model 1 was selected because effects estimated using this approach were expected to be less prone to regression dilution bias. Finally, to predict the geographical variation of anaemia in children aged ≤15 years at unsampled locations we used Model 2. Model 2 was selected because the spatial predictions of parasite prevalence could be used in the generation of anaemia prevalence maps.

Spatial modelling specifications

For the purpose of Bayesian geostatistical modelling parasite prevalence, the child parasite infectious status is

considered a binary outcome variable Y_i which was labelled $Y_i = 1$ for an infected child and 0 for non-infected child. The models used assume a conditional Bernoulli model for the binary outcome variable where the probability of p_j, given the location j of the survey is given by:

$$Y_j \sim Bernoulli\ (p_j)$$

$$log\ (p_j) = \alpha + \sum_{k=1}^{p} \beta_k \times x_j + u_j$$

where Y_j is the infection status of a child in location j, p_j is the probability of an child being infected in location j, is the intercept, x_j is a matrix of environmental covariates, β is a matrix of coefficients for the environmental covariates and u_j is a geostatistical random effect defined by an isotropic powered exponential spatial correlation function:

$$f(d_{ab};\ \phi) = exp\ [-(\phi\ d_{ab})],$$

where d_{ab} are the distances between pairs of points a and b, and ϕ is the rate of decline of spatial correlation per unit of distance.
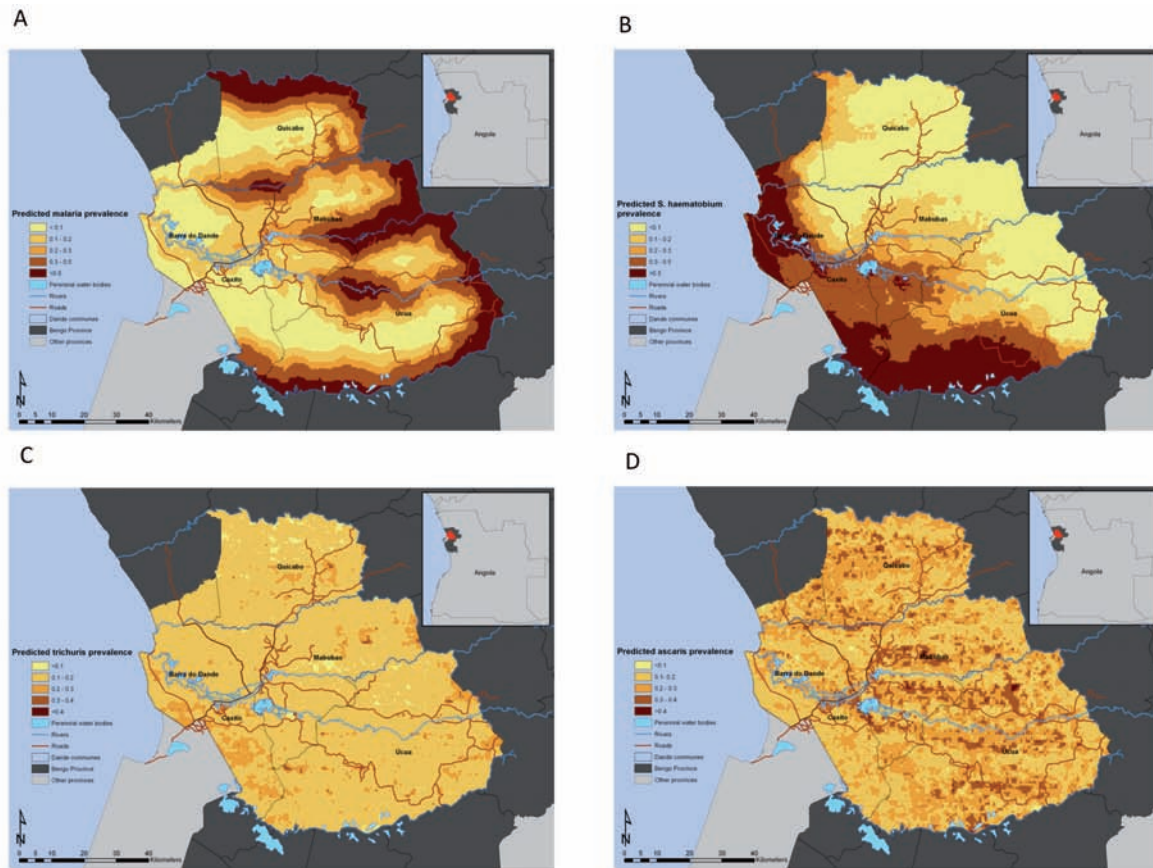
Fig. S3. Predicted spatial distribution of malaria (A), *S. haematobium* (B), *T. trichiura* (C) and *A. lumbricoides* (D) for Dande municipality, Angola.

In the case of modelling prevalence of anaemia, the anaemia status also considered a Bernoulli distributed binary outcome variable Y_i which was labelled $Y_i = 1$ for anaemic and 0 for non-anaemic. In the case of Model 1, x_i is a matrix of individual-level parasite infection covariates. However, in the case of Model 2, x_i was defined by:

$$x_j \sim Beta \ (\alpha_i \ \beta_i)$$

$$\alpha_i = x_i \left(\frac{x_i \ (1 - x_i)}{v_i} - 1 \right)$$

$$\beta_i = (1 - x_i) \left(\frac{x_i \ (1 - x_i)}{v_i} - 1 \right)$$

So the mean predicted parasite prevalence, x_j, was defined by a beta distribution with a shape α_i and a scale parameter β_i; these parameters were estimated using the mean posterior prediction x_i and the mean posterior variance v_i from the parasite prediction models for each helminth survey location. This specification explicitly considers the uncertainty around the

mean parasite predictions (Supplementary Fig. S3) into the spatial geostatistical model of anaemia. Non-informative priors were used for the intercept and the coefficients (normal prior with mean = 0 and precision = 1×10^{-4}). The precision of u_i was given a non-informative gamma distribution.

In all models, a burn-in of 5,000 iterations was allowed, followed by 10,000 iterations where values for the intercept, coefficients and predicted probability of infection at the prediction locations were stored. Diagnostic tests for convergence of the stored variables were undertaken, including visual examination of history and density plots; convergence was successfully achieved after 5,000 iterations.

The predictions of the prevalence of anaemia were made at the nodes of a 0.005×0.005 decimal-degree grid (approximately 500 × 500 m) by interpolating the geostatistical random effect and adding it to the sum of the products of the coefficients for the fixed effects and the values of the fixed effects at each prediction location. Values of predicted prevalence of anaemia at unsampled locations were stored for stunted male children aged ≤15 years. The interpolation of the random

effect was done using the *spatial.unipred* kriging function in WinBUGS; the *spatial.unipred* command implements Bayesian kriging (Thomas, Best et al. 2004) where the values of predicted prevalence at unsampled locations are estimated (interpolated) independently of neighbouring values, as opposed to joint prediction which is conditional on the values of neighbouring unsampled locations. Joint prediction was not considered feasible in this study due to being extremely intensive computationally.

The outputs of Bayesian models, including parameter estimates and spatial prediction at unsampled locations, are distributions termed "posterior distributions". The posterior distributions fully represent uncertainties associated with the parameter estimates. We summarised the posterior distributions in terms of the posterior mean and 95% Bayesian credible interval (CrI), within which the true value occurs with a probability of 95%. The variables age and parasite infection prevalences were standardised by subtracting the mean and dividing by the standard deviation. The resulting regression coefficients for these variables represent the change in terms of standard deviations in prevalence of anaemia that result from a change of one standard deviation in these variables.

Estimation of the population attributable fraction (PAF) of anaemia due to malnutrition and infections

We estimated the *PAF* of anaemia using the standard formula (Rothman, Greenland et al. 2008):

$$PAF_1 = \frac{P_1\,(OR_1 - 1)}{P_1\,(OR_1 - 1) + 1}$$

where P_1 is the mean prevalence of one parasite, and OR_1 is the prevalence-specific odds ratio (OR). The OR for the prevalence of infection with one parasite was estimated by exponentiating the mean posterior estimate of the coefficient (obtained from the spatial prediction model).

References

Rothman KJ, Greenland S, Lash TL, 2008. Modern Epidemiology: Lippincott, Williams, & Wilkins.

Thomas A, Best N, Lunn D, Arnold R, Spiegelhalter D, 2004. GeoBUGS User Manual. Cambridge: Medical Research Council Biostatistics Unit.