Geo-spatial Patterns and Associated Risks of Iron Deficiency and Infection among Young Ghanaian Children: Implications for the Safety of Iron Supplementation in Malaria Endemic Areas

by

Ashley Mariko Aimone

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy Epidemiology

Department of Public Health Sciences
University of Toronto

Geo-spatial patterns and associated risks of iron deficiency and infection among young Ghanaian children: implications for the safety of iron supplementation in malaria endemic areas

Ashley Mariko Aimone

Doctor of Philosophy Epidemiology

Department of Public Health Sciences
University of Toronto

2016

Abstract

Background: The safety and effectiveness of iron supplementation in malaria endemic areas may partly depend on host iron status; however, current methods for assessing iron deficiency risk tend to be confounded by infection and are infeasible to implement at a population level. Determining the geographical patterns of iron status and infection may provide a practical alternative means of identifying high risk populations for whom integrated anaemia and infection control programs are needed.

Objective: Determine the geo-spatial factors associated with iron status and infection risk among 1943 Ghanaian children (6-35 months of age) before and after participating in a randomized iron home-fortification trial.

Methods: Secondary spatial analyses of iron status and infection outcomes were conducted. Iron status was defined as serum ferritin concentration corrected for inflammation (C-reactive protein, CRP) using a regression-based method. Malaria and non-malaria infection outcomes

were defined using four combinations of inflammation (CRP >5 mg/L) and malaria parasitaemia (with and without reported history of fever or concurrent axillary temperature >37.5^0 C). Analyses were performed using a geographical information system (GIS) and generalized linear geostatistical modelling with a Matern spatial correlation function.

Results: After adjusting for demographic characteristics such as age, sex, and maternal education, none of the geo-spatial factors included in the iron status models (including elevation, and distance to a health facility) demonstrated associations at baseline or endline; however, there was significant residual spatial variation across the study area. Conversely, malaria parasitaemia at baseline was associated with greater distance to a health facility and lower elevation. These relationships did not remain at endline, nor when infection was defined using CRP only. Mapping the model outputs showed defined low-risk areas that tended to cluster around villages, particularly near the District centre.

Conclusions: In a malaria endemic area, geographical location may play a role in the risk of iron deficiency and infection among children. Iron home-fortification likely alters the spatial risk profile of malaria and non-malaria infection in this setting, though additional research is needed to confirm the direction of these relationships.

Acknowledgments

There are many people who made this thesis possible, and to whom I would like to extend my utmost gratitude.

My supervisors and committee members: Drs. Donald Cole and Stanley Zlotkin for their unwavering support and unparalleled mentorship, both in doctoral studies and in life. Patrick Brown for teaching me everything I know about spatial statistics and for his incredible patience while I fumbled through the learning process.

The research team in Ghana: Dr. Seth Owusu-Agyei for his time and support in conceiving the project and accessing staff and resources at the Kintampo Health Research Centre. The GIS team for collecting the spatial data for my project and providing continued data management support.

The institutions that I would like to acknowledge include: The University of Toronto, in particular the staff and colleagues in the Departments of Public Health Sciences and Geography for providing learning opportunities, learning support, and technical guidance through coursework and project specific activities. The Canadian Institutes for Health Research (CIHR) for funding my doctoral work, as well as the Bill and Melinda Gates Foundation (through the NIH and NICHD) for funding the original Ghana trial – the foundation of my research.

Finally, I would like to express my enormous gratitude and appreciation for my family and friends who stood by me on this exciting, though sometimes incredibly exhausting, thesis journey. In particular, my husband Jay, for keeping me honest, as well as a roof over our heads and love in our hearts. My parents, Elaine and Peter, for giving me perspective, and purpose, and...well...everything I am. And my coach, Kevin, for keeping me fit and focused, and for being a great mentor and friend.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1
Introduction

1 Introduction

1.1 Background and objectives

Iron deficiency affects more than 1 billion people worldwide and often coexists with infectious diseases such as HIV, TB and malaria, especially in low- and middle-income countries (LMICs) (Stoltfus, Mullany, & Black, 2004). Globally, iron deficiency accounts for approximately 841,000 deaths and 35,057,000 disability-adjusted life years lost (Lopez, Mathers, Ezzati, Jamison, & Murray, 2006; Stoltzfus, 2003), making it the most common micronutrient deficiency in the world. Public health initiatives to reduce iron deficiency are needed; however recent concerns about the safety and effectiveness of iron supplementation in settings of high infectious disease prevalence have slowed these efforts. These concerns reached a breaking point in 2006 when the results of a large randomized trial, conducted in Pemba, Zanzibar, indicated that providing iron supplements to young children living in a malaria endemic area may increase the risk of severe morbidity and mortality from malaria and other infections, particularly among those who are iron *replete* (Sazawal & et al., 2006). At the time, there were a multitude of proposed biological mechanisms thought to play a role in the risk relationships observed. For example, when iron supplements are provided to children who are not iron deficient, there may be an accumulation of free iron in the blood stream, which is essential for multiplication of bacteria and parasites (Weinberg, 1999) The authors concluded that, in malaria endemic areas, the identification and treatment of iron deficiency could be a safer approach than universal iron supplementation.

The findings in Pemba created a global reaction of fear and uncertainty regarding the safety of iron supplements, particularly in sub-Saharan Africa where malaria is a leading cause of infection-related morbidity and mortality among children. Part of this fear was related to the recommendation at the time that, in areas where malaria transmission is high and infectious

diseases prevalent, iron supplementation (drops, syrup or tablets) be targeted only to children who are anaemic and at risk of iron deficiency, and receiving concurrent protection from malaria and other infectious diseases (WHO/UNICEF, 2006) ;. In other words, before initiating iron treatment, not only should there be strategies in place for infection prevention, but also differentiation of iron *depletion* versus iron *repletion*. While haemoglobin concentration is routinely used for assessing iron status at a population level (e.g. by assuming that 50% of all anemia cases are due to iron deficiency) (WHO, 2015; WHO/CDC, 2007); in a malaria endemic area this approach may lead to overestimation of the true prevalence of nutritional iron deficiency (due to the haemolytic effect of malaria parasitaemia) (Balarajan, Ramakrishnan, Ozaltin, Shankar, & Subramanian, 2011), and thus put children at risk of the adverse effects of supplementing with iron during concurrent infection. On the other hand, commonly used indicators of iron status, such as serum ferritin, are confounded by infection due to the up-regulating effect of inflammation on acute phase proteins (D.I. Thurnham & McCabe, 2012). Despite this limitation, serum ferritin is generally recommended as the single best biomarker for assessing nutritional iron status and evaluating the response to iron interventions, as it correlates closely with body iron stores (Mei et al., 2005; WHO, 2011a). Unlike haemoglobin, however, measuring serum ferritin concentration at the population level is costly and labour intensive, and thus generally not feasible or practical in low- or middle-income countries.

In 2007, the research community responded to the global reaction to the Pemba trial through a WHO-commissioned consultation, which led to the initiation of the "Iron and Malaria Project". The Iron and Malaria Project consisted of a targeted research agenda with the goal of generating evidence and informing policy and practice in three core areas (WHO, 2006):

1. To identify plausible *mechanisms* to explain the adverse effects of iron supplementation on infection risk;
2. To assess *biomarkers* that can accurately and reliably measure nutrition iron deficiency in areas of endemic infection; and
3. To evaluate the safety and effectiveness of *interventions* to prevent and treat iron deficiency.

One of the funded studies under the Iron and Malaria Project was a cluster-randomized trial investigating the impact of home fortification with micronutrient powders (MNPs) on malaria incidence among children in rural Ghana (Zlotkin et al., 2013). Briefly, the children were randomized at the compound (residence) level to receive a daily dose of MNPs with or without 12.5 mg iron (plus other micronutrients) for 5 months during the rainy season (when malaria transmission is typically higher). It was hypothesized that providing iron as a micronutrient powder would be a safer alternative to supplements in terms of malaria infection risk. Unlike the Pemba trial, all participants in the Ghana trial also received bed nets at enrolment, as well as malaria treatment throughout the trial when indicated. The intention-to-treat analysis showed that the incidence of clinical malaria was significantly lower in the Iron-group compared to the No-iron group (294 versus 344 total episodes, or 68.8 versus 78.8 episodes/100 child-years, respectively; RR 0.87, 95% C.I. 0.78-0.96), though this difference was no longer statistically significant after adjusting for baseline iron deficiency and anemia status. Despite lower malaria rates, hospital admission rates were higher in the Iron group compared to the No-iron group, for reasons that could not be elucidated given the available data (Zlotkin et al., 2013). It was concluded that, in a malaria endemic setting where appropriate malaria control measures are in place, daily use of iron-containing MNPs may be safe with respect to malaria incidence; however, additional precaution may be needed in terms of potential interactions with host iron status, and the impact of iron on susceptibility to other (non-malaria) infections (Prentice, Verhoef, & Cerami, 2013). These issues emphasize the need for accurate and reliable biomarkers or indicators that can differentiate between iron *deficiency* and iron *sufficiency* among populations that also have a high risk of malaria and other infections.

Geographical factors could fulfill the need for an indicator that provides insight into the dynamics and distribution of iron status among children in malaria endemic areas, that is not confounded by infection or inflammation, and that is feasible to measure at a population level. Enhancing our knowledge of the geographical patterns and spatial factors associated with iron status could inform treatment needs (Schur, Vounatsou, & Utzinger, 2012), and help implementers and policy makers in LMICs to determine where health services should be targeted in order to make efficient use of limited resources. Geo-spatial data are often publicly

available, which also improves the access to and comparability of population-level trends across regional and national borders.

Geographical information systems (GIS) and spatial analysis methods have been used in several areas of health research, such as disease surveillance and risk analysis, assessment of health system access, and health system planning (Masters et al., 2013; Nykiforuk & Flaman, 2008; Root et al., 2014; Song, Zhu, Mao, Li, & An, 2013). Pairing cartographical GIS outputs (e.g. dot density figures or choropleth maps) with quantitative spatial analyses can also be used to investigate more complex relationships involving both space and time, such as those associated with intervention evaluations. For example, Bhatt and colleagues used Bayesian hierarchical spatiotemporal modelling to quantify *P. falciparum* malaria prevalence and disease incidence across sub-Saharan Africa from 2000 to 2015, as well as the impact of major malaria intervention coverage during that time (Bhatt et al., 2015). Using data from 27,573 georeferenced population clusters, along with temporally varying environmental and socio-demographic variables, they estimated that 663 million (95% credible interval 542, 753 million) clinical cases had been averted since the year 2000, and that insecticide-treated nets were the most widespread intervention and the largest contributor to this reduction (68% of cases averted) (Bhatt et al., 2015).

In terms of nutrition research, however, the investigation of geo-spatial factors associated with iron deficiency risk among children in LMICs has been limited, and almost entirely restricted to the use of haemoglobin concentration or anaemia status as a primary outcome (Aimone, Perumal, & Cole, 2013). In a systematic review, we found 5 publications that compared anaemia prevalence among children 0-5y of age across geographic regions or assessed the association of anaemia with spatial factors in low- and middle-income countries (Aimone et al., 2013). Considering the risks of providing iron supplements to children with high infection exposure (e.g. in malaria endemic areas) who may be anaemic though not iron deficient, it is important to be able to differentiate these states in order for the benefits of iron deficiency control programs to outweigh their risks and costs. I hypothesized that spatial analyses could provide a

complementary means of identifying paediatric populations at risk of nutritional iron deficiency (and sufficiency) in settings with a high malaria burden, that is non-invasive and more feasible compared to measuring biomarkers at a population level.

To test this hypothesis, I conducted a secondary analysis of biochemical and demographic data from the Ghana trial, with the addition of primary geographical data that were collected after trial completion. The primary objective of these analyses was to use GIS and model-based geostatistics to explore the utility of geo-spatial factors (or "geo-indicators") in the assessment of iron deficiency and infection risk among children living in rural Ghana. As illustrated in the conceptual frameworks below (**Figures 1.1 and 1.2),** the key constructs of this objective, in its simplest terms, surround the relationship between health and the environment. More specifically, I am interested in determining whether spatial relationships play a role in the variation in iron deficiency risk (**Figure 1.1**) or infection risk (**Figure 1.2**), and thus whether geographical location could be an additional tool for improving the effectiveness of program targeting and treatment response. To explore these constructs with the least amount of bias, I considered all explanatory variables to be potential direct or indirect predictors of iron deficiency or infection risk, though the spatial variables were of primary interest. In addition to their role as predictors, certain spatial and non-spatial variables were also postulated be effect modifiers (e.g. land cover type, child iron status, child infection status).

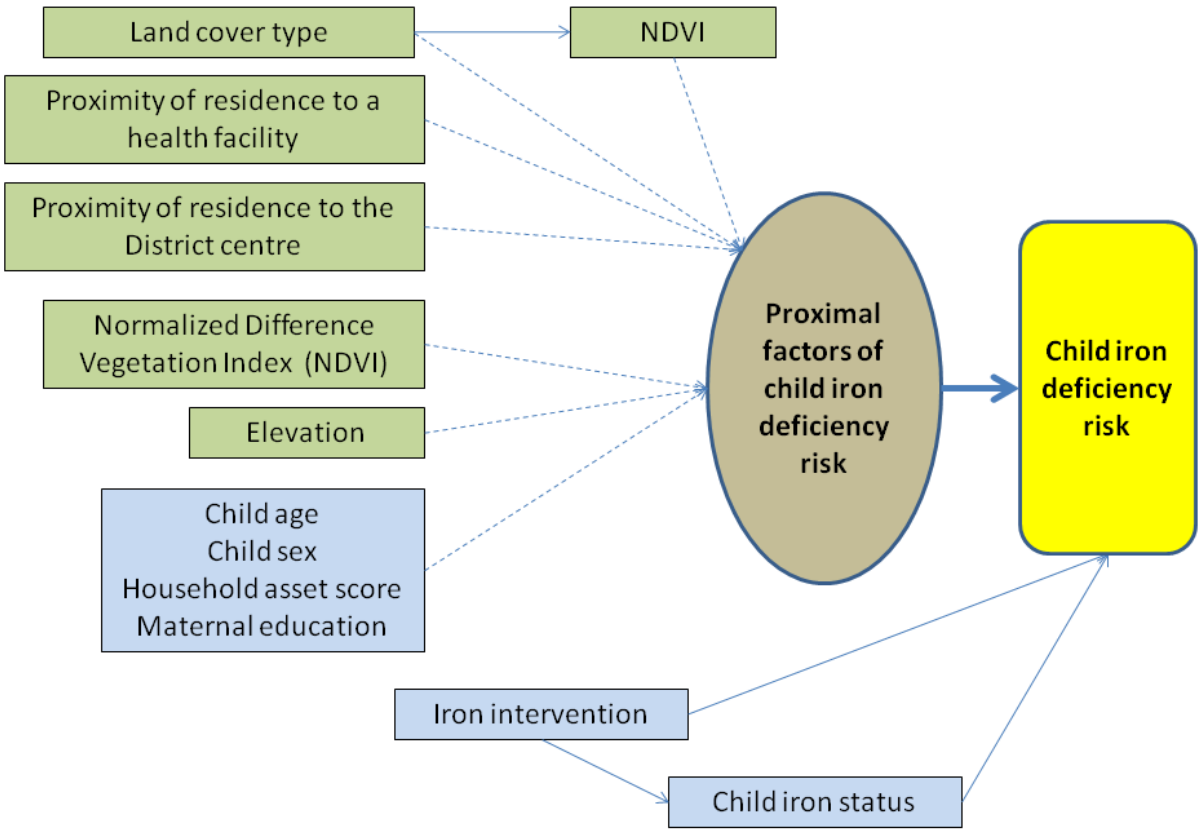


Figure 1.1: Conceptual framework illustrating the role of independent variables (boxes) in terms of their hypothesized direct (solid arrows) or indirect (dashed arrows) relationships with iron deficiency risk. Green boxes represent variables generated from primary spatial data, while blue boxes represent variables generated from secondary demographic and biochemical data.

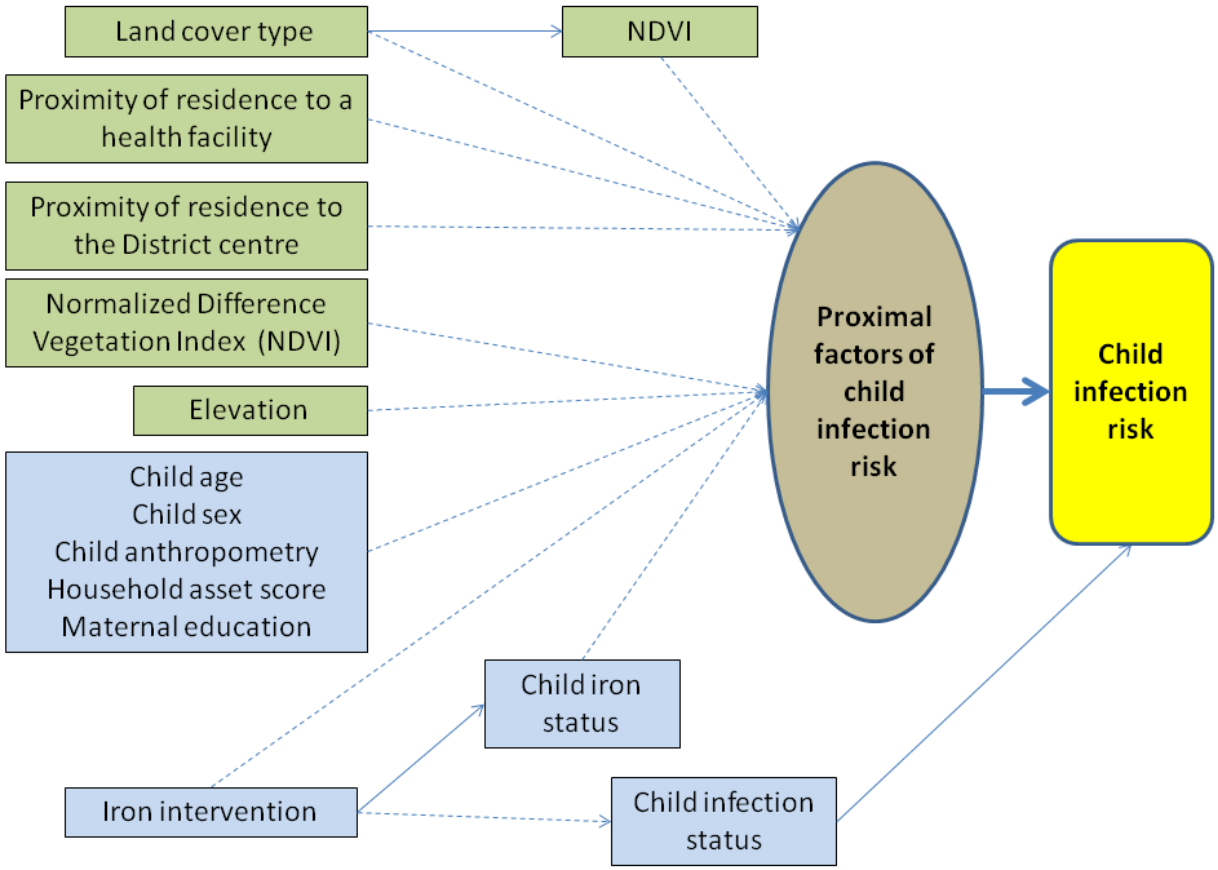


Figure 1.2: Conceptual framework illustrating the role of independent variables (boxes) in terms of their hypothesized direct (solid arrows) or indirect (dashed arrows) relationships with infection risk. Green boxes represent variables generated from primary spatial data, while blue boxes represent variables generated from secondary demographic and biochemical data.

In order to explore the constructs and variable relationships pertaining to the primary objective, the analyses were broken down into *three specific aims*:

1. To determine the geo-spatial patterns of *iron status* (ferritin concentration) among young Ghanaian children before and after participating a randomized trial of iron home-fortification;
2. To determine the geo-spatial factors associated with the variation in *infection risk* among young children in rural Ghana; and
3. To determine the impact of iron home-fortification with micronutrient powders on the geo-spatial patterns of *infection risk* among young children in rural Ghana

In addition to knowledge enhancement, spatial analyses could also contribute to the goals of the Iron and Malaria Project (in terms of biomarkers, mechanisms, and interventions) by adding another dimension to our understanding of the distribution and dynamics of iron deficiency and infection risk that may not be apparent through evidence generated strictly from basic science studies and clinical trials (**Figure 1.3)**. More specifically, in Aim 1, geo-indicators could provide additional insight into the assessment of biomarkers that can accurately and reliably measure nutritional iron status. Similarly, in Aim 2, the analysis of baseline data may indicate whether geo-indicators can contribute to the assessment of biomarkers for measuring infection status. Lastly, in Aim 3, geo-indicators may add to our knowledge and understanding of the safety and effectiveness of iron interventions to treat and prevent iron deficiency in malaria endemic areas; as well as the mechanisms to explain the adverse effects of an iron intervention on infection risk. Another potential advantage of geographical information systems (GIS) and spatial analysis methods is through knowledge translation and exchange, as geographical maps are often easily interpreted by a lay audience and across political and cultural boundaries. Therefore, such visual representations of research results (e.g. geographical risk profiles) can facilitate their communication and thus provide an additional tool to aid in the advancement or implementation of relevant programs or policies.

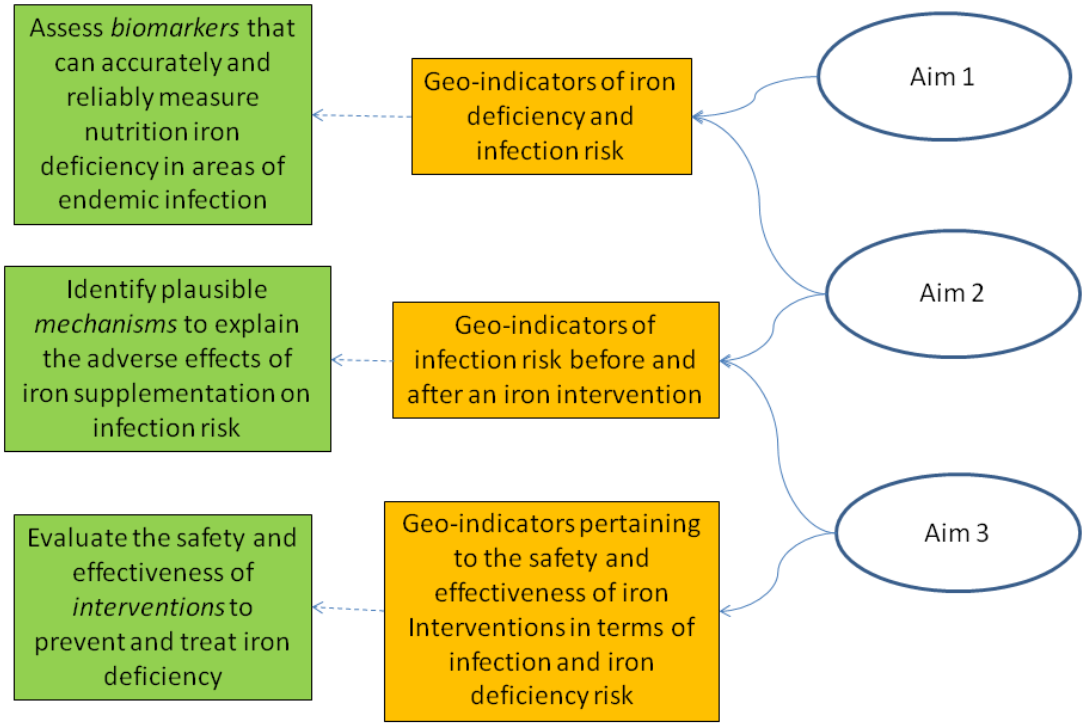


Figure 1.3: Contribution map illustrating the hypothesized role of each thesis aim (ovals) in terms of their hypothesized contribution (orange boxes) towards the main research pillars of the Iron and Malaria Project (green boxes). Solid arrows represent direct contribution, while dashed arrows represent indirect contribution.

1.2 Chapter map

The subsequent chapters will start with a literature review (Chapter 2) and a description of the methods used (Chapter 3). Following this are detailed accounts of the analyses associated with each aim, formatted as individual manuscripts (Chapters 4, 5 & 6). The final chapter includes an overall discussion of the findings by aim, linked to key current research themes on iron and malaria, as well as the contribution of this thesis and directions for application and future work (Chapter 7). References for Chapters 1-3 and 7 can be found at the end of Chapter 7, while references pertaining to each aim are at the end of their respective Chapters (4-6).

1.3 Author contributions

I conceived this thesis project at the conclusion of the Ghana trial, for which I assisted in writing the original proposal, was involved as a coordinator, and also contributed to the final publication. I coordinated the *post-hoc* acquisition of geographical data from the Ghana study area and, with support from my supervisory committee, designed and conducted the secondary analysis and interpreted the results. In terms of overall contribution, I foresee that the findings from these analyses will introduce a new perspective on the intersection between nutritional status and infectious disease, particularly in the context of iron nutrition and malaria endemicity.

Chapter 2
Literature Review

2 Literature Review

The following literature review begins with an overview of the role of iron in human health, and the prevalence and impact of iron deficiency in childhood (2.1). Options for preventing or correcting iron deficiency are also discussed, as well as the efficacy and effectiveness of each as a population health intervention. The subsequent section introduces the iron and malaria 'problem', and how mounting evidence on the interaction between iron nutrition and infectious disease and the adverse effects of iron supplementation has influenced the global health research community (2.2). Section 2.3 focuses on one of the main facets of the iron and malaria problem, which pertains to the challenges of measuring and interpreting biomarkers of nutritional iron status in the context of high infection burden. The final section (2.4) explores the potential utility of geographical information systems (GIS) and spatial analysis methods as a means of assessing or predicting iron deficiency risk in a malaria endemic area that can overcome some of the challenges of traditional biomarker-based approaches.

2.1 Iron and child health

Iron plays an important role in human metabolism, neurological function, haematology, the immune system, and physical development (Beard, 2001; WHO/UNICEF/UNU, 2001). Nutritional iron deficiency is a major global health concern that affects approximately 2 billion people worldwide (Stoltfus et al., 2004). According to the Global Burden of Disease Study 2013, iron deficiency anaemia was estimated to account for 36,663.5 (24,371.0-53,084.7) years lived with a disability (YLDs) in 2013, ranking third among the top ten causes of YLDs

globally, and second in Sub-Saharan Africa(Collaborators, 2015). Children under five years of age make up a large proportion of the iron deficiency burden, accounting for 20,854 deaths (2.2 million disability adjusted life years lost) per year (Black et al., 2008). Children have high iron requirements for red blood cell production (erythropoiesis) during the first few months after birth, which are normally met through stores that were acquired during the last trimester of pregnancy. After 6 months of age, a child's iron requirements for growth and development must be met through dietary sources. Children in low- and middle-income countries are at risk of iron deficiency because they are often born with low iron stores and fed unfortified plant-based complementary foods that are poor sources of dietary iron (FAO/WHO, 2005). If iron deficiency is left untreated, it can lead to anaemia and its functional consequences such as impaired cognitive performance and resistance to infection, poor growth, adverse pregnancy outcomes, and reduced work capacity and productivity as adults, leading to socioeconomic regression (WHO/NHD, 2001). Recent estimates of anaemia prevalence suggest that around 800 million women and children worldwide are affected, with the highest prevalence among children less than 5 years of age (43%, 95% C.I. 37-47%), particularly in central and west Africa (71%, 95% C.I. 67-74%) (Stevens et al., 2013). Although iron deficiency tends to be the largest contributor to anaemia prevalence (approximately 42% of all cases), in African countries the percent of cases attributable to iron deficiency is lower than the global average (32%), which suggests other potential causes, such as malaria and genetic haemoglobinopathies like sickle cell and thalassemia (Stevens et al., 2013).

Options for treating or preventing iron deficiency include: 1) increasing the diversity of the diet (with a focus on sources of iron with high bioavailability); 2) general fortification of the food supply with iron; 3) targeted fortification of commercially available complementary foods, such as infant cereals; 4) bio-fortification through genetic modification of staple crops (e.g. rice, beans, cassava); 5) iron supplementation (pills, syrups or drops); and 6) home fortification of prepared foods at the point of consumption with iron-containing micronutrient powders. For children living in developing countries, the first three options may not be feasible due to cost-related barriers to availability and accessibility. Bio-fortification and general fortification of the food supply is also not appropriate for children, since the level of fortification is often not

adequate to meet their high daily iron requirements. Iron supplements, particularly in the form of syrups and drops, tend to have low acceptability and compliance due to gastrointestinal side effects, metallic taste, teeth staining, and the need for literacy to measure individual doses. Furthermore, there is evidence to suggest that providing iron supplements to children at risk of exposure to infection, particularly malaria, can lead to adverse events including hospitalization and death (Sazawal & et al., 2006).

Home fortification is a concept that lies between targeted fortification and individual supplementation, which involves adding specialized products, such as micronutrient powders (MNPs) to foods prepared in the home (HF-TAG, 2011). MNPs are blends of vitamins and minerals in powder form administered by sprinkling them onto food immediately before it is consumed. MNPs enable families without access to commercially fortified foods to add essential vitamins and minerals directly to traditional foods prepared in the home. They are inexpensive to produce, have no special storage requirements, and are simple to use even by those who are not literate. Compared to iron syrups, they also tend to be more acceptable to mothers and children because the microencapsulation of the iron prevents staining of the teeth, and strong metallic aftertastes (Zlotkin, Arthur, Antwi, & Yeung, 2001). In 2011, the WHO published guidelines recommending the use of home fortification (or point-of-use fortification) with iron-containing MNPs to improve iron status and reduce anaemia among infants and children 6-23 months of age (WHO, 2011b). A subsequent global assessment of home fortification interventions by UNICEF and the Centers for Disease Control and Prevention (CDC) identified 42 countries with implemented or planned interventions using MNPs with an expected reach of 14.1 million people (Jefferds, Irizarry, Timmer, & Tripp, 2013).

2.2 Iron and infection

The challenge of iron deficiency is not just in the magnitude of the problem, but also in the inherent properties of iron itself as potential pro-oxidant (Halliwell & Gutteridge, 1989) or

pathogen enhancer, as well as the complex biological systems needed to maintain iron homeostasis (Andrews & Schmidt, 2007). While iron deficiency can have a negative impact on the immune response to infection, "anaemia of infection" can also result as part of an innate immune response whereby iron is withheld from invading pathogens (through sequestration and reduced absorption) in order to prevent microbial growth and proliferation (Ganz & Nemeth, 2009; Greenberg & Ashenbrucker, 1947; Murray, Murray, Murray, & Murray, 1978; Wander, Shell-Duncan, & McDade, 2009). So the fundamental biological questions that have been raised include: 1) whether host iron status has an impact on the immune iron-withholding defense mechanism, and 2) whether providing iron supplements (in an effort to treat or prevent iron deficiency) can disrupt or enhance this immune response. Despite ongoing concerns regarding the interactions between iron status, iron supplementation, and susceptibility to infection (Oppenheimer, 2001), the importance of iron in child health could not be overlooked, and thus the World Health Organization (WHO) recommended universal iron supplementation for infants and children in areas where the prevalence of anaemia exceeded 40% (WHO/UNICEF/UNU, 2001). It was not until 2006 that this recommendation was revised when the results of a large randomized trial, conducted in Pemba, Zanzibar, indicated that providing iron supplements to young children living in a malaria endemic area may increase the risk of severe morbidity and mortality from malaria and other infections, particularly among those who are iron replete (Sazawal & et al., 2006). Considering the global prevalence of malaria at the time was estimated as 350-500 million clinical disease episodes per year (WHO/UNICEF), the findings in Pemba caused great concern in the global health community and prompted the WHO and United Nations Children's Fund (UNICEF) to release a joint statement recommending that in malaria endemic areas, iron supplementation (drops, syrup or tablets) be targeted only to children who are anaemic and at risk of iron deficiency (WHO/UNICEF, 2006).

The global response to the Pemba trial created fear and uncertainty regarding the safety of iron supplements, particularly in sub-Saharan Africa where malaria is a leading cause of infection-related morbidity and mortality among children (WHO, 2014). In addition to revising its recommendations, the WHO also commissioned a consultation (Lyon, France) to develop evidence-based recommendations for iron administration in malaria-endemic regions. As

mentioned previously (Section 1.1), the results of the Lyon consultation included the identification of three core research areas pertaining to *mechanisms, biomarkers, and interventions* (WHO, 2006). In 2007, a collaboration was initiated between the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) of the U.S. National Institutes of Health (NIH), the Bill & Melinda Gates Foundation (BMGF), and the WHO to address the issues raised by the Pemba trial and the research priorities identified through the Lyon consultation. The collaboration fostered the launch of the "Iron and Malaria Project", which consisted of a targeted research agenda and knowledge translation initiatives with oversight by a Steering Committee, a Technical Working Group (TWG), and a Research Review Committee. The targeted research agenda was guided by a technical report, generated by the Iron and Malaria TWG, which included a review and assessment of the current state of evidence with respect to the three focal research areas (mechanisms, biomarkers, and interventions), as well as recommendations for further research in these areas (D. Raiten, Namaste, & Brabin, 2009).

The *mechanisms* thought to be responsible for the harmful effects of iron supplementation included: (i) The production of plasma non-transferrin bound iron (NTBI) with increased amounts of absorbed iron, (ii) the effects of increased amounts of iron in the gastrointestinal tract on intestinal structural integrity and microflora, and (iii) the complex immune effects of iron interventions (G.M. Brittenham, 2012). Iron supplements are often administered in high doses in a post-prandial state resulting in a rate of iron influx that exceeds the rate of iron uptake by transferrin, an iron transport protein. Current evidence from human studies suggests that a higher rate of iron absorption compared to the capacity to bind the absorbed iron (with transferrin) can result in the formation of NTBI or "free iron" that is accessible to and can enhance the growth of pathogenic microorganisms (D. Raiten et al., 2009). It was further hypothesized that NTBI increased the intensity of malaria infection by increasing the sequestration of malaria-infected erythrocytes in the capillaries of the brain and intestine, leading to cerebral malaria and increased permeability of the intestinal barrier to pathogens (R. F. Hurrell, 2011). The immune function of gut bacteria could then be further impaired through

frequent high doses of supplemental iron by stimulating the growth of pathogenic bacteria in the intestine (R. Hurrell, 2010).

While increased NTBI was confirmed in iron supplementation studies, it had not been shown with consumption of iron fortified foods (G.M. Brittenham, 2012; G. M. Brittenham et al., 2014; R. Hurrell, 2010). Therefore, the NTBI hypothesis formed the basis of the "Ghana trial", one of the *intervention* studies within the Iron and Malaria Project. The Ghana trial was a cluster-randomized intervention trial investigating the impact of home fortification with iron-containing micronutrient powders on malaria incidence among young children (Zlotkin et al., 2013). The trial was conducted in a rural area of Ghana, which is a malaria endemic country with an estimated 3·2 million cases in 2008, most caused by *P. falciparum* (WHO, 2009). I participated in the Ghana trial as a project coordinator, assisting in the oversight of recruiting and following 1958 infants and young children (6-35 months of age) from 22 villages in the Wenchi and Tain Districts of the Brong-Ahafo Region. Enrolled participants were randomized at the compound level to the Iron or No-iron group. Children in the Iron group received a daily dose of MNP with 12.5 mg iron (plus 30 mg ascorbic acid, 400 μg vitamin A, and 5 mg zinc) (IOM, 1997-2001) for 5 months during the rainy season (April-November) followed by a 1-month follow up period. Those in the No-iron group received a similar MNP without iron. Compound-level randomization was used primarily to avoid cross-contamination between intervention groups through food sharing among households. This randomization scheme also facilitated tracking of participants, particularly those who moved within or outside of the trial area during the intervention or post-intervention period.

It was hypothesized that providing iron as a micronutrient powder would be a safer alternative to iron pills or drops in terms of malaria incidence because MNPs are administered with food and the iron is microencapsulated, resulting in a potentially reduced peak labile pool of free iron in the circulation that is available to invading parasites. All participants in the Ghana Study also received bed nets at enrolment, as well as malaria treatment throughout the trial when indicated. The intention-to-treat analysis showed that the incidence of clinical malaria was significantly

lower in the Iron-group compared to the No-iron group during the intervention period (294 versus 344 total episodes, or 68.8 versus 78.8 episodes/100 child-years, respectively; RR 0.87, 95% C.I. 0.78-0.96), as well as the follow-up period (338 versus 291 total episodes, or 79.1 versus 89.7 episodes/100 child-years, respectively; RR 0.87, 95% C.I. 0.79-0.97). The differences were no longer statistically significant after adjusting for baseline iron deficiency (ferritin <12 mcg/L after excluding samples with CRP >8mg/L) and anemia status (Hb <10g/dL) (Zlotkin et al., 2013). Despite lower malaria rates, hospital admissions were higher in the Iron group compared to the No-iron group. Although data availability did not allow for a detailed assessment of the reason for the higher number of hospitalizations, it was hypothesized that the iron intervention may have interacted with host iron status to create a "transient susceptibility" to non-malaria infections (Prentice et al., 2013).

Considering the results of the Ghana Study, as well as supporting literature at the time (Ojukwu, Okebe, Yahav, & Paul, 2009; Okebe, Yahav, Shbita, & Paul, 2011), it could be concluded that, in a malaria endemic setting where appropriate malaria control measures are in place, daily use of iron-containing MNPs may be safe with respect to malaria incidence; however, additional precaution may be needed in terms of potential interactions with host iron status, as well as the impact of iron on a child's susceptibility to other (non-malaria) infections. The latter issues bring us back to need for accurate and reliable *biomarkers* or bio-indicators for identifying iron deficiency risk among populations that are also at risk of malaria and other infections.

2.3 Biomarkers for assessing iron status

As suggested by the results of the Pemba trial, as well as the subsequent WHO recommendation, in order to ensure the safety of iron supplementation in a malaria endemic area, iron should only be provided to those who are iron *deficient* (WHO/UNICEF, 2006). While haemoglobin concentration is widely used to screen for anaemia at a population level, it is not sensitive or specific enough to identify anaemia due to nutritional iron deficiency,

particularly in malaria endemic areas where haemolysis of parasitized erythrocytes and other red blood cell polymorphisms are common causes of low haemoglobin (D. J. Raiten, Namasté, & Brabin, 2011). On the other hand, serum ferritin concentration correlates closely with body iron stores, and is generally recognized as the single best biomarker for assessing iron status and evaluating the response to iron interventions (Mei et al., 2005; WHO, 2011a). Unlike haemoglobin measurements, however, determining serum ferritin concentration is a resource intensive process that often requires large blood samples, specialized analytical equipment, and trained personnel. Therefore, assessing iron status at a population level using ferritin may not be feasible or practical, particularly in low- and middle-income countries where the prevalence of malaria and other infections also tends to be high.

The measurement challenge is further complicated by the role of ferritin as an acute phase protein, which is up-regulated in the presence of inflammation or infection (D. I. Thurnham, Mburu, Mwaniki, & De Wagt, 2005). To assist in the interpretation of serum ferritin in situations where the prevalence of infection or inflammatory disease is high, the WHO recommends raising the cut-off used to define iron deficiency (e.g. to 30 µg/L), or excluding individuals with elevated concentrations of inflammatory markers such as C-reactive protein (CRP), alpha1-antichymotrypsin (ACT), or alpha1-acid-glycoprotein (AGP) (WHO, 2011a; Worwood, 2007). The success of these methods has been inconsistent, in part because of the variation across inflammatory markers in terms of the type of inflammation and stage of disease they represent. For example, CRP is a positive acute phase protein (APP) that represents the early phases of the inflammatory response (incubation and early convalescence), as it becomes elevated approximately 4-6 hours after infection onset, reaches its peak concentration within 24-48 hours, and typically falls off quickly as clinical symptoms resolve. On the other hand, AGP is a slower reacting positive APP that reaches its peak concentration within 4-5 days, and thus represents the later phases of the inflammatory response (early and late convalescence) (D. J. Raiten et al., 2015; D.I. Thurnham & McCabe, 2012; Worwood, 2007). The rise in plasma ferritin concentration tends to parallel that of CRP; however it tends to remain elevated when CRP concentration drop off, after which it more closely resembles the concentration curve of AGP (D.I. Thurnham & McCabe, 2012). A practical approach may be to ignore serum ferritin values when CRP or AGP is elevated; however, this may lead to the exclusion of a large amount of data and thus create biased prevalence estimates (D. Raiten et al., 2009).

Thurnham and colleagues proposed a method for calculating a correction factor for serum ferritin based on the values of CRP and AGP (D. I. Thurnham et al., 2010). The advantage of the correction factor method is that it attempts to account for the stage of infection (e.g. incubation vs. early convalescence); however, it is limited by necessitating the use of pre-determine cut-offs for CRP and AGP, which can vary across the literature partly due to the detection limits of analytical equipment used. This cut-off issue has recently been addressed by the research group for the Biomarkers Reflecting Inflammation and Nutrition Determinants of Anemia (BRINDA) study (P.S. Suchdev et al., 2016). The BRINDA group proposed a new method for correcting ferritin for inflammation using regression modelling, which was developed using 17 datasets, representing 15 countries and over 25,000 children less than 5 years of age (P.S. Suchdev et al., 2016). Similar to Thurnham's method, the regression method can be used to correct ferritin for CRP or AGP or both simultaneously; however, it does not require pre-determined cut-offs for either inflammatory marker and therefore better accounts for the linear relationship between inflammation and ferritin (P.S. Suchdev et al., 2016).

Despite all of the adjustment strategies for interpreting ferritin values, there is still a recognized need for valid and reliable biomarkers or bio-indicators for assessing iron deficiency risk in the context of high infection exposure (D.J. Raiten & Combs, 2015). Investigations in this area are ongoing, and include new methodologies that incorporate other biomarkers that are less influenced by inflammation, such as serum transferrin receptor (sTfR) and zinc protoporphyrin (ZPP) (Menendez et al., 2001; Thomas, Koenig, Lightsey, & Green, 1977). The evidence to date supporting the utility of these biomarkers has been inconsistent, and thus more extensive evaluations are required before specific recommendations for use can be made (D. Raiten et al., 2009). Another promising biomarker that has appeared more recently in the literature is hepcidin, a regulatory hormone that controls the transport of iron into the plasma via the transport protein ferroportin (Donovan et al., 2005). When iron status is adequate, hepcidin concentration is increased, which restricts entry of iron into plasma by degrading ferroportin and inhibiting iron absorption (Ganz, 2005; Nemeth et al., 2004). Hepcidin is also increased during

the inflammatory response to infection, which acts to restrict microbial growth by reducing access to iron in the plasma (Wander et al., 2009). Therefore, hepcidin could potentially be used as a biomarker of iron "readiness", indicating when iron support is needed (i.e. when iron status is low) and/or when iron absorption and utilization can be maximized (i.e. when inflammation is absent). While further research is needed to confirm the validity of new methods for correcting ferritin, and/or determine the clinical application of emerging biomarkers like hepcidin (van Santen, de Mast, Swinkels, & van der Ven, 2011), an overarching issue that impedes the utility all biomarker-related approaches is the need for blood sampling and laboratory-based analyses.

2.4 Geo-indicators

The need for biomarkers and other molecular-level methods for assessing nutrient status or infection risk may never be eliminated entirely; however, it could be reduced (or optimized) if there were other macro-level measures that provide complementary means of identifying populations that should be prioritized for further biological assessment. Furthermore, the conceptual gap between a biomarker and higher-level public health indicators can be expansive, and make inferences about risks and benefits difficult to generate or translate into clear guidance for implementers and policy-makers (D.J. Raiten & Combs, 2015). Therefore, a macro-level indicator should also provide relevant contextual information for interpreting a biomarker and its implications for public health action (D.J. Raiten & Combs, 2015). An attribute that could fulfill this role, is the lived environment. In other words, *where* individuals live is an important aspect of health that could be linked to specific biomarkers, and thus add another dimension to our understanding of how they contribute to population health dynamics. In the case of this project, determining the link between geographical characteristics and the variation in biomarkers of iron deficiency or infection risk could provide a means of pre-*screening* for prevention and treatment targeting that is non-invasive and less costly to measure compared to the biomarkers themselves.

When it comes to defining indicators, Raiten and Combs (2015) emphasize the importance of language and terminology. According to these authors, the distinction between an indicator and biomarker reflects, not only the level at which it is measured, but also the expectations about what the information means and how it should be used (D.J. Raiten & Combs, 2015). They further emphasize that the evidence generated from investigations of biomarker and indicator relationships should be translatable to a broad community of potential users with varying degrees of technical expertise (D.J. Raiten & Combs, 2015). Following Raiten and Combs' suggested terminology that differentiates biomarkers (sensitive and specific measures of nutrient exposure, status, and function) from bio-indicators (sentinel measures of functional change associated with nutritional status or intervention), I have proposed the term *geo-indicator* in reference to any geographical characteristic or spatial attribute at a point location that is also linked to the risk of a health outcome at that location. Geo-indicators can provide a context-specific complementary means of identifying populations that are at risk of iron deficiency or infection, and thus should be prioritized for further assessment using biomarkers before delivering an iron intervention. Additional advantages of geo-indicators is that they are less invasive or costly to obtain compared to biomarkers, and geo-spatial data can be easily collected with minimal training (e.g. using hand-held geographical positioning system units, or through open online databases).

2.5 GIS and spatial analysis in health research

2.5.1 Geographical information systems

A geographic information system (GIS) is used to capture, manage, analyze, and display geographically referenced data (ESRI, 2011). Applying a GIS to health research can provide a different way of looking at ecological health data that is easy to interpret by a lay audience and across disciplines. In low- and middle-income countries especially, assessing geographical patterns of health outcomes and risks can inform treatment needs, and thus help implementers

and policy makers to determine where and how medications or supplements should be targeted and delivered in order to make the most efficient use of limited resources (Masters et al., 2013; Nykiforuk & Flaman, 2008; Root et al., 2014; Song et al., 2013).

The first recognized use of GIS in public health dates back to the 19th century when a cholera outbreak in Soho, London was attributed to contaminated water from a local water pump (Buechner, Constantine, & Gjelsvik, 2004). Since this early example, GIS applications have become much more prominent in the public health literature. In 2008, a technical report published by the Centre for Health Promotion Studies at the University of Alberta (Alberta, Canada) identified 621 articles that included the use of GIS in public health investigations under four major themes: Disease surveillance (n=227); risk analysis (n=189); health access and planning (n=138); and community profiling (n-115) (Nykiforuk & Flaman, 2008). The breadth of articles included in the report demonstrates the versatility of GIS as an enabling technology that allows users from various fields of expertise to integrate spatial data into a range of applications. The use of GIS has also expanded globally, partly due to the development of open-access platforms, as well as initiatives by the United Nations (UN) and other government organizations (e.g. IDRC, WHO and CDC) to increase the accessibility and availability of GIS software for low- and middle-income countries (ESRI, 2011).

2.5.2 Spatial analysis

There are a multitude of applications that demonstrate the power of GIS and its ability to further our understanding of environmental factors that influence health-related behaviours or practices. For example, cartographical GIS outputs (e.g. dot density figures or choropleth maps) pair well with quantitative (multivariate) spatial analyses to investigate more complex spatial relationships (Nykiforuk & Flaman, 2008). In the case of paediatric nutrition research, the application of GIS and spatial analysis to investigate geo-indicators associated with iron deficiency or anaemia has been limited.

In a systematic review, we found 5 publications (including peer-reviewed articles and WHO reports) that compared anaemia prevalence among children 0-5y of age across geographic regions or assessed the association of anaemia with spatial factors in low- and middle-income countries (Aimone et al., 2013). For example, in north and sub-Saharan Africa, inverse associations were found between the prevalence of moderate anaemia and the level of urbanization (Brink et al., 1983; Mainardi, 2012). Using geographically weighted regression modelling, Mainardi et al. reported a significant positive association between access to a water source (defined by median travel time) and the prevalence of moderate anaemia (haemoglobin 7-9.9 g/dl) among children <5 years of age across 20 sub-Saharan African countries ($p<0.01$) (Mainardi, 2012). One study also demonstrated a relationship between the spatial distribution of anaemia prevalence and malaria endemicity. Snow et al. compared the health outcomes of young children (0-4 years of age) from two East African communities (Kilifi, Kenya and Ifakara, Tanzania) with markedly different levels of malaria transmission and found that the community with higher parasitaemia also had a higher prevalence of severe anaemia (Snow et al., 1994).

A major drawback of these studies was the use of anaemia as an outcome measure, which is not a specific bio-indicator of nutritional iron deficiency, particularly in sub-Saharan Africa where malaria, haemoglobinopathies, thalassemias, and other conditions leading to anaemia of infection are common. Soares Magalhaes and colleagues attempted to account for non-nutritional anaemia by mapping the geographical distribution of mean haemoglobin (Hb) and anaemia risk among pre-school children (1-4 years of age) in West Africa (Burkina Faso, Ghana, and Mali) using spatial models that adjusted for predicted *Plasmodium falciparum* parasitaemia rate, as well as *Shistosoma haematobium* and hookworm infections (Soares Magalhães & Clements, 2011). They found that moderate anaemia (Hb 70-99 g/L) prevalence exceeded 60% in all study countries, and estimated that 14.9%, 3.7%, and 4.2% of all anaemia cases could be attributed to malaria, *S. Haematobium*, and hookworm infections, respectively (Soares Magalhães & Clements, 2011). In a similar analysis of children (≤15 years of age) in

Angola, Soares Magalhaes et al. identified clusters of high anaemia risk, and found that malnutrition and parasitic infections were important contributors to the spatial variation in anaemia risk (Soares Magalhães et al., 2013). Without a more specific biomarker of iron status, such as serum ferritin, the risk maps developed by Soares Magalhaes et al. may have excluded children who were iron deficient without anaemia, leading to potential underestimation of the depth and geographical breadth of micronutrient deficiency risk across the study areas. A strength of these analyses, however, was the incorporation of infection risk into the anaemia models, which highlighted the magnitude of influence that infection exposure can have on anaemia prevalence, and thus the importance of considering these factors when developing effective interventions to treat iron deficiency in areas where the burden of infection is high.

2.5.3 Geostatistics and the Bayesian approach

The analysis of spatial patterns and processes is a relatively specialized statistical discipline that evolved from applications in ecology (e.g. forestry) to a wide range of research disciplines, where much progress has been made in the description, identification and estimation of spatial stochastic processes (Cressie, 1993). *Geostatistics* refers to a statistical methodology that deals with data consisting of observations at discrete sets of sampling locations within a spatial region, where each observed value is statistically related to an underlying continuous spatial phenomenon (P. Diggle & Ribeiro, 2007). Spatial data are defined as realizations of a stochastic process indexed by space: $Y(s) = [y(s), s \epsilon D]$ (Blangiardo, Cameletti, Baio, & Rue, 2013), which can be represented by a collection of observations, $y = [y(s_1),...,y(s_n)]$ where the set $(s_1,...,s_n)$ indicates the spatial units at which measurements were taken (Blangiardo et al., 2013). If D is a continuous surface (e.g. air pollutant measurements) then the problem can be specified as spatially continuous (Gelfand, Diggle, M., & Guttorp, 2010). If D is a collection of countable spatial units (e.g. number of malaria cases per village) then the problem can be specified as a discrete random process (Gelfand et al., 2010). The parameters in geostatistical data can be defined as a function of some hyper-parameters ψ associated with a suitable prior distribution $p(\psi)$ (Blangiardo et al., 2013). A spatially structured covariance matrix is often specified by the

(isotropic) Matern spatial covariance function (Cressie, 1993), as the Matern family is a flexible class of covariance functions that can cover a range of spatial fields. Also included in the spatial model is a parameter for measuring the degree of smoothness of the process (the order of the Bessel function), and a scaling parameter that is related to the range (i.e. the distance at which spatial correlation is close to zero) (Cressie, 1993). Examples of the types of problems that can be addressed using geostatistical methods include:

1. Description of the elevation of a malaria endemic region through a continuous elevation map that interpolates the data, as well as the true underlying surface (Davis, 1972); then relating this to the spatial variation in malaria prevalence across the region to determine if (and how) they are associated.
2. Development of a predictive model for the variation in the prevalence of iron deficiency among villages in a malaria endemic area as a function of measured explanatory variables (such as elevation, bed net use and greenness of surrounding vegetation), while accounting for unexplained variation within or between villages (P. Diggle, Moyeed, Rowlingson, & Thomson, 2002).

For geo-spatial analyses, Bayesian methods are attractive alternatives to likelihood-based (frequentist) inference, as the computations required for frequentist inference are intractable when data are not normally distributed (Fong, Rue, & Wakefield, 2010). Since its inception as Bayes' theorem (by Thomas Bayes, 1701-1761), Bayesianism became known as an approach for interpreting probability as a sort of partial belief, rather than a frequency, allowing the application of probability to a wide variety of propositions without the need for a reference class (Bellhouse, 2004). Since the early days of Thomas Bayes, Bayesian methods have developed greatly and are now used in many research areas including epidemiology (Greenland, 2006) and the social sciences (Jackman, 2009). According to Bayesian methodology, uncertainty is described by probability distributions, and thus there is no real distinction between observable data and unobservable parameters (also known as random quantities). Before any new data are observed, the uncertainty about the realized value of the unobserved parameters is described by a *prior* distribution, which is estimated given the current state of information. The prior is then

combined with a new data model to derive the *posterior* distribution using the inferential process (Blangiardo et al., 2013). The specification of a prior distribution allows for the formal inclusion of information from previous studies or expert opinion. The posterior distribution, in turn, allows for easy estimation of a (posterior) probability that is more intuitive and interpretable than the frequentist p-value. Additional advantages of using Bayesian methods are the ease of specifying a hierarchical structure on the data, making predictions for new observations, and imputing missing data (Rue & Martino, 2007).

In terms of its application, Bayesian methods offer an effective approach for analyzing epidemiological data, which often have a spatial and/or temporal structure (Dunson, 2001). For example, if an outcome or risk factor is observed at point locations then a geostatistical model is applied (P. Diggle & Ribeiro, 2007), which can be specified using Bayesian methods by extending the concept of hierarchical structure to account for similarities based on area-level (neighborhood) or point-reference (distance) data (Blangiardo et al., 2013). Diggle et al. (2002) developed a spatial GLMM with Bayesian inference and a Markov chain Monte Carlo (MCMC) implementation to describe the spatial variation in the prevalence of malaria among children in the Gambia (P. Diggle et al., 2002). The analysis methodology was adapted from the model-based geostatistics framework described by Diggle and others in 1998 (P. J. Diggle, Moyeed, & Tawn, 1998), with the purpose of extending a previous analysis by Thomson et al. (1999) using the same Gambian dataset (Thomson et al., 1999). Thomson et al. used a marginal regression method (generalized estimating equations) to account for extrabinomial variation in a model of malaria prevalence with child-level, village-level, and spatial covariates (Thomson et al., 1999). According to Diggle et al., accommodating for extrabinomial variation in a marginal regression analysis will treat spatial variation as a ‘nuisance’ effect, and thus does not fully explain any underlying spatial variation or allow a smooth interpolation of residual spatial variation (P. Diggle et al., 2002). They argued that if strong extrabinomial variation exists, Bayesian model-based geostastical methods could assess whether this variation is spatially structured. Evidence of spatial structure means that, at least part of, the explanation for the variation must be environmental. Otherwise, there may be unmeasured characteristics (e.g. shared village-level factors) that are inducing non-spatial extrabinomial variation (P. Diggle et al., 2002).

2.5.4 Integrated nested Laplace approximation (INLA)

Since Diggle et al.'s work in 2002, several other studies have used Bayesian geo-statistical methods to produce maps of malaria risk, in sub-Saharan Africa (Gemperli et al., 2004; Giardina et al., 2012; Gosoniu, Veta, & Vounatsou, 2010; Raso et al., 2012) and worldwide (Hay et al., 2009) that could be used to identify regions most in need of targeted interventions to reduce malaria burden. Another aspect that these studies have in common is the use of a Markov chain Monte Carlo (MCMC) implementation method, which has been the traditional implementation method for GLMMs (Brooks, Gelman, Jones, & Meng, 2011; Rue & Martino, 2009), though also represents a potential downside of Bayesian analyses. While flexible and able to deal with many types of data, MCMC methods involve computationally and time-intensive simulations, which can limit the feasibility of increased model complexity and data dimensions. Faster and more computationally convenient alternatives to MCMC, such as integrated nested Laplace approximations (INLA) have been recently proposed (Fong et al., 2010; Held, Schrodle, & Rue, 2010; Rue & Martino, 2009). The INLA method exploits the assumptions of a spatial model to produce numerical approximations to posteriors of interest, based on the *Laplace approximation* (Tierney & Kadane, 1986). INLA first explores the marginal joint posterior for the hyperparameters to locate the mode; then a grid search is performed and produces a set of points with a corresponding set of weights to give the approximation to the posterior distribution (Rue & Martino, 2009). The key to the INLA method is to construct *nested* approximations, which makes Laplace approximations very accurate when applied to *latent Gaussian models* (Rue & Martino, 2009). Latent Gaussian models are a subclass of structured additive regression models where the latent field is Gaussian (a Gaussian Markov random field, GMRF) with a sparse precision matrix (Rue & Held, 2005), and a dependent variable that is assumed to belong to the exponential family (i.e. non-Gaussian) (Rue & Martino, 2009). INLA can be used in a variety of applications, such as disease mapping (Schrodle & Held, 2011), time trend analyses (Riebler, Held, & Rue, 2012), and diagnostic test studies (Paul, Riebler, Bachmann, Rue, & Held, 2010).

Three recent studies used INLA to model malaria risk in Africa. Alegana et al. used a zero-inflated conditional-autoregressive model with Bayesian inference and INLA to investigate the spatial and temporal variation of malaria incidence across 273 health facilities in northern Namibia in 2009 (Alegana et al., 2013). After adjusting for test positivity rates and health facility utilization, the mean annual incidence was predicted to be 13 cases per 1000 population with the highest incidence in constituencies bordering Angola and Zambia (Alegana et al., 2013). The analysis output also included smoothed maps, which showed trends in malaria incidence across the northern regions that could be used as a baseline for monitoring disease elimination targets (Alegana et al., 2013). In Burkina Faso, Samadoulougou and colleagues used Bayesian methods with INLA to map parasitaemia risk in 2010 among 6,102 children less than 5 years of age, as well as examine the individual-, household-, and community-level characteristics and climatic/environmental factors associated with malaria infection (Samadoulougou et al., 2014). The results of the analysis indicated that malaria risk was associated with child age, household wealth index, maternal education, health facility attendance, normalized difference vegetation index, monthly rainfall, and population density (Samadoulougou et al., 2014). Similar to Alegana et al., the authors concluded that the analysis results and corresponding risk maps could be used by relevant decision-makers to identify priority areas where malaria control efforts should be enhanced (Samadoulougou et al., 2014).

Lastly, Bhatt and colleagues used Bayesian hierarchical spatiotemporal modelling to quantify *P. falciparum* malaria prevalence and disease incidence across sub-Saharan Africa from 2000 to 2015, as well as the impact of major malaria intervention coverage during that time (Bhatt et al., 2015). A geostatistical model was constructed using data from 27,573 georeferenced population clusters, along with temporally varying environmental and socio-demographic variables. They estimated that 663 million (95% credible interval 542, 753 million) clinical cases had been averted since the year 2000, and that insecticide-treated nets were the most widespread intervention and the largest contributor to this reduction (68% of cases averted) (Bhatt et al., 2015). The authors not only demonstrated the utility of geostatistical methods in tracking the progress of public health initiatives, but also suggested that an even more robust surveillance

system would be an important part of continued malaria control efforts; especially if prevalence continues to decline to the point of elimination (Bhatt et al., 2015).

2.6 Summary

Iron plays an important role in human development, and the prevention and treatment of iron deficiency in children should be prioritized globally. Recent evidence has indicated that caution is needed when addressing iron deficiency in settings with a high prevalence of infectious disease, particularly malaria. Findings from the literature have also suggested that, in order to ensure the safety of iron interventions in malaria endemic areas, it is important to be able to accurately assess nutritional iron deficiency; however, commonly used biomarkers are difficult to measure and interpret in these settings due to their high resource costs and the confounding effect of infection. I proposed that GIS and spatial analysis could provide a complementary means of identifying or predicting the location of paediatric populations at risk of iron deficiency that is more feasible than measuring biomarkers and less influenced by infection status. Spatial analysis methods have been used in various areas of health research such as disease surveillance and risk analysis; though have not been extensively applied to child nutrition, particularly in relation to iron deficiency. The following chapter provides a detailed account of how I investigated the utility of spatial methods in iron and infection research by using GIS and model-based geostatistics to explore the geographical patterns and environmental factors associated with iron status and infection risk among children living in a malaria endemic area.

Chapter 3
Methods

3 Methods

3.1 Data sources and collection

3.1.1 Clinical trial population

The data for the secondary analyses were obtained from a previously conducted cluster-randomized clinical trial investigating the impact of home fortification with powdered micronutrients with/without iron on the incidence of clinical malaria among Ghanaian children (Zlotkin et al., 2013). The study population from the Ghana trial included infants and young children (6-35 months of age) from 22 villages in the Wenchi and Tain Districts of Brong-Ahafo, which is a rural region in the middle of Ghana. The Brong-Ahafo region is the second largest in the country (39,558 km^2) with the third lowest population density (58/km^2 in 2010) (GSS, 2012) (**Figure 3.1**). The region lies primarily in the forest zone, which is a major cocoa- and timber-producing area. The northern part of the region lies in the savannah zone, a large grains- and tubers-producing area (Ghana, 2016). The population of Wenchi and Tain (153,633 in 2010) is representative of the general Ghanaian population, particularly the rural regions. The number of children younger than 5 years of age in Wenchi and Tain (11,215 in 2010) represents approximately 0.3% of the total pre-school aged child population in Ghana (11,215/3,533,000). There are a total of 99 villages across the two Districts, consisting of 8,548 compounds (residence consisting of one or more households), for an average of 86 compounds per village. However, the range is quite large (up to 2,100 compounds per village) due to variations in village size and/or population density. The trial area covered approximately 3,800 km^2, which included 22 villages and 1,552 compounds. A village was eligible for inclusion in the study if the inhabiting households had at least one child between 6 and 35 months of age.

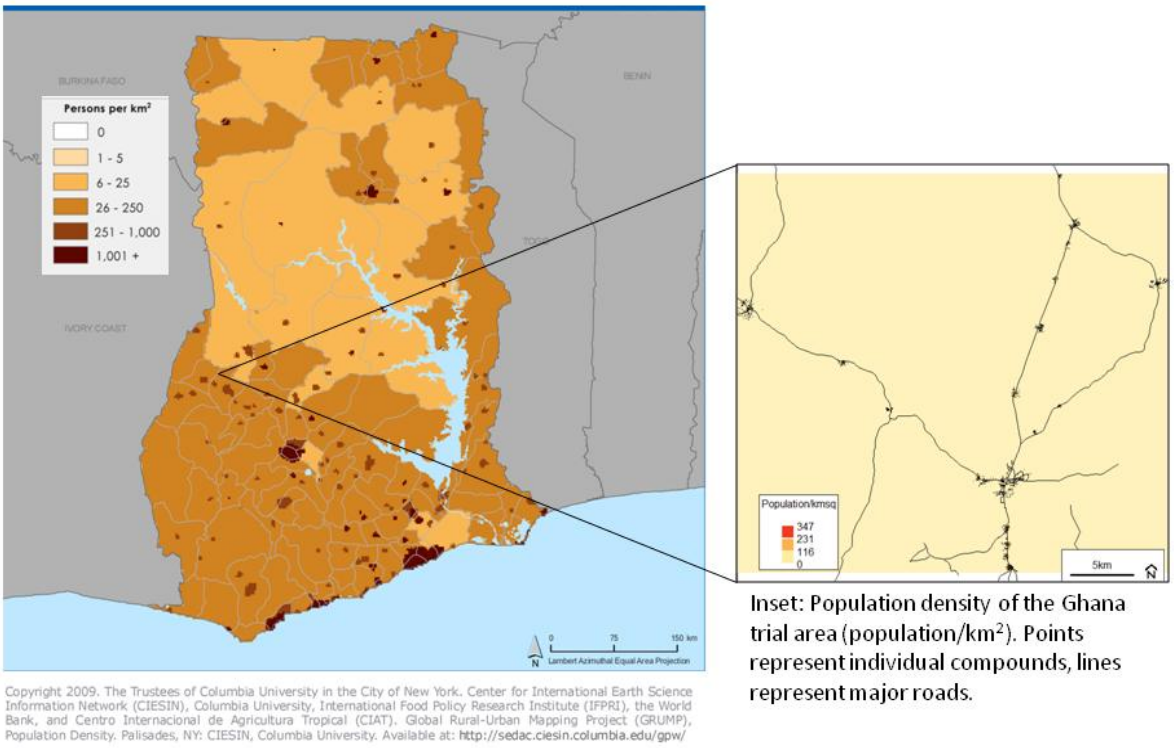


Figure 3.1: Population density of Ghana (left) and the Ghana trial area (right) in 2000. The trial area (inset) was located in the Wenchi and Tain Districts of the Brong-Ahafo Region. Country map adapted from the Socioeconomic Data and Applications Centre (SEDAC, 2009) – available at http://sedac.ciesin.columbia.edu.

All compounds in the two Districts had been previously enumerated, including the number of children per compound, for a recently completed randomized trial investigating the impact of weekly vitamin A supplementation in on pregnancy-related mortality (Hurt et al., 2013). These databases were used to identify potentially eligible participants, who were then screened, beginning with compounds nearest to the Kintampo Health Research Centre (where the study team was based), then moving to adjacent villages along the main road network. Eligible children were aged 6-35 months, eating solid foods, and living in the study area for at least the following six months. Exclusion criteria included severe anaemia (haemoglobin <7.0 g/dL), severe malnutrition (weight-for-length z-score <-3·0), receipt of iron supplements within the past 6 months, or chronic illness (e.g. congenital abnormalities). Out of 2220 children screened, 200 met the exclusion criteria and 62 were absent for enrolment, leaving 1958 who were randomized at the compound (cluster) level to the Iron or No-iron group (**Figure 3.2**). A

compound consisted of one or more households living in the same residence. The number of enrolled clusters was 1552 giving an average cluster size of 1.3 participants (range 1-5). At the end of the 5-month intervention, 1815 children (n=900 in the Iron group and n=915 in the No-iron group) were followed-up for endline data collection, including a blood sample and anthropometric measures (7.3% attrition). Participants with incomplete follow-up had either moved away (n=39), died (n=5), were temporarily absent (n=89), or absent for unknown reasons (n=10).

Cluster randomization was used primarily to avoid cross contamination between intervention groups through food sharing, and to facilitate tracking of participants throughout the study period. Although randomization at the village level may have been preferable, it also would have greatly reduced the statistical power of the analyses, given that there were only 22 villages (compared to over 1500 compounds). Furthermore, the original sample size for the Ghana trial was calculated based on individual-level units of observation, and thus randomizing at the compound level allowed for minimal loss in power while optimizing intervention compliance. Previous micronutrient powder (MNP) intervention studies conducted in a similar area of Ghana (Zlotkin et al., 2001; Zlotkin et al., 2003) have used compound-level randomization with minimal risk of cross-contamination. To further offset this risk, the weekly follow-up surveys in the Ghana trial included questions about adherence, including whether the MNPs were lost, sold, or shared with other family members. Over the entire intervention period, there were only four accounts of reported MNP sharing, which were equally distributed between the Iron and No-iron groups.

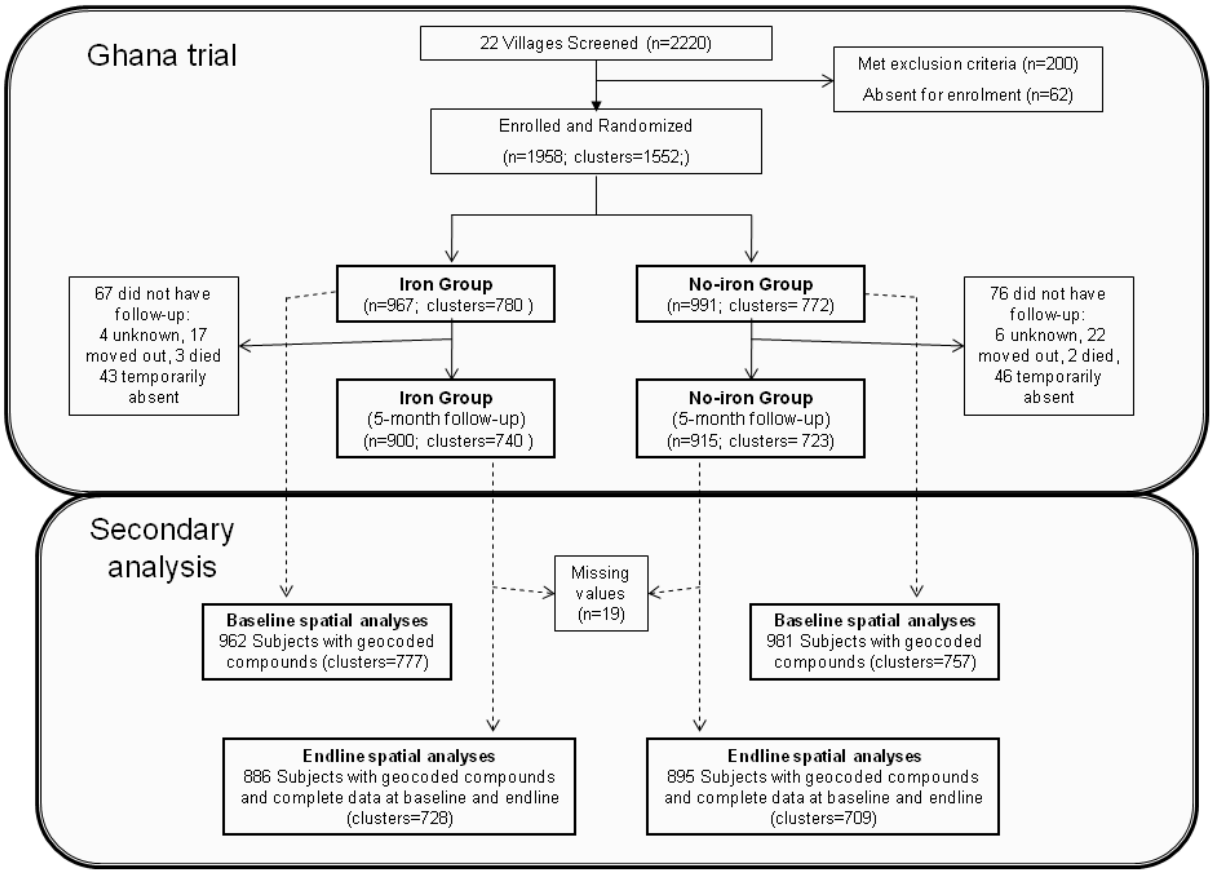


Figure 3.2: Study flow for the Ghana trial (top section) and secondary analyses (bottom section). Out of the 1958 participants from the Ghana trial, a total of 1943 with geocoded compounds were included in the baseline secondary spatial analyses (13 compounds were untraceable, corresponding to 15 participants not included in the secondary analyses). The endline spatial analyses included a total of 1781 observation, representing trial participants with geocoded compounds who provided blood samples at the end of the intervention period (5-month follow-up), and after removing 19 observations due to missing baseline or endline ferritin values.

3.1.2 Secondary clinical trial data

The variables of interest for the current secondary analyses were those pertaining to iron and infection status, including serum ferritin, haemoglobin, plasma C-reactive protein (CRP), and malaria parasite density. All biomarkers were measured at baseline and the end of the 5-month intervention period in the Kintampo Health Research Centre laboratory facility. Serum ferritin

was measured using an enzyme immunoassay (Spectro Ferritin S-22, Ramco Laboratories Inc., Stafford, USA), and haemoglobin was determined as part of a complete blood count using a hematology auto-analyzer (Horiba ABX Micros 60-OT-CT-OS-CS, Montpellier, France). CRP was measured using an immunoturbidimetric assay (QuickRead CRP, Orion Diagnostica, Espoo, Finland), and malaria parasite density was determined using microscopy (thin and thick smears). Malaria screening was also performed on a weekly basis throughout the intervention period using a rapid diagnostic antigen test (Paracheck Pf®). If a child had a fever within the past 48 hours or an axillary temperature >37.5^{0}C, a blood sample was drawn and analyzed in the field via antigen test, and in the lab using microscopy. Antigen test results were used to determine treatment needs, and parasite density was combined with fever information to calculate clinical malaria incidence (episode counts). Additional details on these laboratory methods can be found in **Appendix A**. Demographic and nutrition-related information was collected at baseline at the household and individual levels, including household assets, maternal education, feeding practices, and child body weight and length (**Appendix B**). Z-scores for weight-for-length, length-for-age were calculated using the WHO Child Growth Standards (WHO/UNICEF, 2009).

3.1.3 Geographic primary and secondary data

For the geostatistical analyses, primary spatial data were collected *post hoc* from the study area, including global positioning system (GPS) coordinates for study compounds, villages, surrounding health facilities, schools, churches, local water sources (e.g. wells), major markets, and road networks. Only 13 compounds (0.8%) were untraceable, resulting in a nearly complete analytic sample of 1943 participants with geocoded compounds (99.2% of the original sample size of 1958 participants). As part of my PhD work, I coordinated the collection of these data by a field team from the Kintampo Health Research Centre in October 2012, approximately two years after completion of the clinical trial. The field team measured all coordinates using standard hand-held GPS receiver units; set to the WGS84 datum and UTM zone 30N projection system (EPSG code: 32630). Coordinate data were imported into a geographical information system (GIS), and stored in a GIS software-compatible format (e.g. shape files, DBF tables).

Satellite-derived elevation data (SRTM DEM Version 3) were downloaded as a Georeferenced Tagged Image File Format (GeoTIFF) file from the United States Geological Survey (USGS) (USGS, 2012). The latter is a government agency for natural sciences that provides information on the global environment, drawing on data such as the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) data sets. The data file had a resolution of 3 arc-seconds (approximately 90 meters) (**Figure 3.3**). Normalized Difference Vegetation Index (NDVI) data were downloaded from the Land Processes Distributed Active Archive Center (LPDAAC) (USGS, 2014). The NDVI data was produced by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument (product MOD13Q1), which uses blue, red, and near-infrared reflectance to determine vegetation indices for 16-day intervals with 250-meter spatial resolution (**Figure 3.4**). The NDVI value for a given raster cell (pixel) is calculated from the visible (VIS) and near-infrared (NIR) light reflected by vegetation [NDVI = (NIR-VIS)/(NIR+VIS)], which results in a number that ranges from negative one (-1.0) to positive one (+1.0) (Weier & Herring, 2000). The MODIS land cover type product (MOD12Q1) was downloaded from worldgrids.org and consisted of 17 land cover classes (defined by the International Geosphere Biosphere Programme or IGBP), including natural vegetation (11 classes), developed and mosaic land (3 classes), and non-vegetation (3 classes). Land cover data were derived from yearly output from the Terra and Aqua MODIS instruments at a resolution of 500 meters (Hengl, 2014) (**Figure 3.5**).

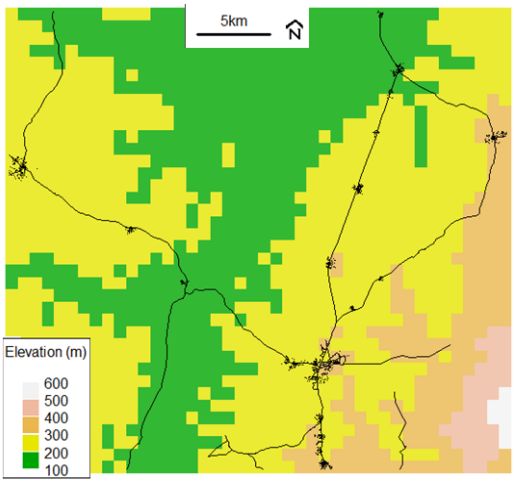


Figure 3.3: Elevation (meters) of the Ghana trial area. Points represent individual trial compounds. Lines represent major roads.

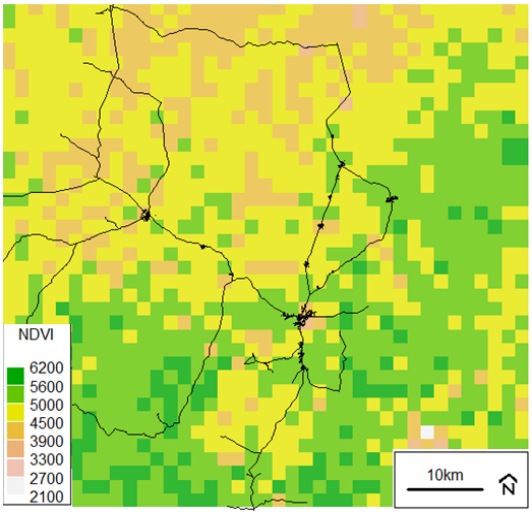


Figure 3.4: Normalized difference vegetation index (NDVI) of the Ghana trial area (averaged for 2010). Points represent individual trial compounds. Lines represent major roads.

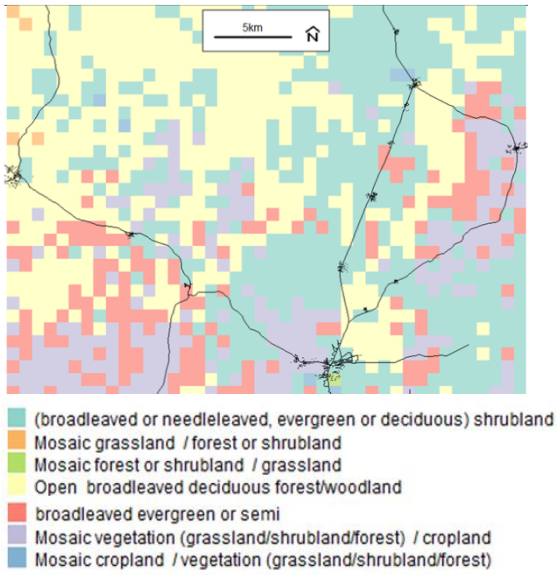


Figure 3.5: Land cover type of the Ghana trial area. Points represent individual trial compounds. Lines represent major roads.

3.1.4 Ethics

Approval for the original clinical trial was obtained from the Kintampo Health Research Centre (KHRC) Institutional Ethics Committee, the Ghana Health Service (GHS) Ethical Review Committee, the Hospital for Sick Children Research Ethics Board, and the Food and Drugs Board of Ghana. Approval for conducting the secondary analysis of trial data, as well as the collection and primary analysis of geographical data was obtained from the Hospital for Sick Children and University of Toronto Health Sciences Research Ethics Boards.

3.2 Independent variable descriptions

Baseline age in months was calculated using the reported date of birth and trial enrolment date. Similar to the linear regression analyses, the age variable was included in all spatial models with a change point at 24 months, as this was the closest half-year to the mean age of those children who were no longer receiving breast milk (mean=26.8 months ± 5.8, n=746). A change point of 24 months has also been used in other studies of iron deficiency and anaemia in children (Adish, Esrey, Gyorkos, & Johns, 1999; Jonker et al., 2012). Household asset score was generated using a principal component analysis of 6 economic indicators, including farm ownership, size and type of crops grown (e.g. cash crop for sale in the market or for household consumption), type of toilet facility, and house ownership (e.g. own vs. rent). For descriptive purposes, asset score was dichotomized at the median. Maternal education was included as a dichotomous variable (no education versus any education). Distance to the nearest health facility was included as an indicator of access to the health care system. Distance to the District centre (Wenchi Town) was included as an indicator of access to other community-level factors that can affect health (e.g. local markets, schools, pharmacies, and employment opportunities) (Shirayama, Phompida, & Shibuya, 2009). All distance types were measured `as the bird files' (straight-line or Euclidean) in kilometers using the Near Table tool in ArcMap (ArcGIS 10.2, Environmental Systems Resource Institute, Redlands, California).

All satellite-derived data were downloaded as global datasets and cropped according to the geographical boundaries of the study area. Elevation had a range of 116-530 meters, and values were centered by subtracting 250 before including the raster in the analyses. Centering the cell values of a raster can heighten the visual contrast between low- and high-valued areas (P. Diggle & Ribeiro, 2007), as well as facilitate the interpretation of spatial interactions. Land cover type (LC) was a discrete categorical variable consisting of 3 values representing woody savannahs (LC=8, n=21/1943 observations), urban and built up land (LC=13, n=243/1943 observations), and cropland/natural vegetation mosaic (LC=14, n=1679/1943 observations). For modelling purposes, the largest category (cropland/natural vegetation mosaic) was used as the reference.

Since the Brong-Ahafo Region of Ghana has a tropical climate with a yearly double rainfall maxima separated by a long dry season, the vegetation (or greenness of the area) has the propensity to change over the short term in response to water availability. Therefore, NDVI was considered to be a proxy for soil moisture (Samadoulougou et al., 2014), and thus an indication of the tendency for water to collect and create breeding grounds for Anopheles mosquitoes. While the rainy season in northern Ghana typically falls between March and November, the timing, duration, and intensity of rainfall can vary slightly from year to year, which also influences the variation in malaria transmission. In an effort to reduce the impact of this variation and create a more generaliseable variable, all NDVI data that were available for Ghana during the year that the study was conducted (2010) were averaged in a single raster file. Averaged NDVI values for the trial area in 2010 had a range of 0.22-0.62. Two NDVI*LC interaction terms were created by centering the average NDVI raster values (NDVI / 1000 – 4), and adjusting the resolution of the LC raster to match that of the NDVI raster. The NDVI raster was then used to mask the LC raster except where LC=8 (these cells were given a value of zero). This created a new raster that could be used to investigate whether the association between the dependent variable and vegetation (or soil moisture) varied across areas with or without a woody savannah land cover type. The same method was applied for the NDVI*LC interaction term with LC values of 13 (urban and built-up land).

3.3 Dependent variable exploration

3.3.1 Linear modeling and simulation analyses

The objective of the linear modeling and simulation analyses was to determine the suitability of using ferritin concentration to define iron status by exploring its relationship with potential confounding factors, such as anaemia (haemoglobin concentration) and infection (CRP concentration, parasitaemia with or without fever). In other words, these analyses were a sort of "proof of concept" in terms of confirming that haemoglobin (Hb) was *not* an appropriate proxy for iron status in the Ghana trial population; and demonstrating that infection status could in fact explain a large amount of variation in ferritin concentration. The linear models also provided an

opportunity to compare ferritin correction methods and infection definitions with respect to the impact on the predictive ability of Hb and infection biomarkers, respectively.

A first set of linear regression models of baseline ferritin concentration was created with or without Hb as an independent variable, along with age and sex (Model L1.1). The distribution of ferritin concentration was positively skewed ($D=0.24$, $p<0.01$ for the Kolmogorov-Smirnov test of normality), and thus log-transformed values were used in the analyses. Age in months was calculated using the reported date of birth and study enrolment date with a change point at 24 months. In order to account for the effect of inflammation, ferritin concentration was modelled in two additional ways: in Model L1.2 ferritin was uncorrected and CRP was included as a covariate with a quadratic transformation. Generalized additive models (GAM) were used to examine the relationship between CRP and (log) ferritin in order to determine the most appropriate transformation for the CRP covariate. In Model L1.3, ferritin was corrected for inflammation (CRP) using the regression method described by Suchdev et al. (P.S. Suchdev et al., 2016). Briefly, the regression method consists of the following analytical steps:

1. Categorize the inflammation marker (CRP) into (internal) deciles and extract the value of the lowest decile (α), which will be subtracted from the value of CRP in the final step.
2. Regress CRP against ferritin (log-transformed) and extract the beta coefficient (β). If the beta coefficient is positive then the effect of CRP on ferritin is subtracted from the observed ferritin value. Conversely, if the beta coefficient is negative, the effect of CRP on ferritin is 'added back' to the measured ferritin value.
3. Calculate ferritin corrected for CRP-inflammation using the following formula (where SF=measured serum ferritin): $SF\_corrected = SF + (CRP-\alpha)*(-\beta)$

A second set of linear regression models of baseline serum ferritin concentration was created with or without infection status as an independent variable, along with age, sex, and Hb concentration. As in the first set of models, the ferritin dependent variable was log-transformed, and age had a change point at 24 months. Infection status was included in the models in four

ways: 1) CRP>5 mg/L and/or malaria parasitaemia (Model L2.1); 2) CRP>5 mg/L without malaria parasitaemia (Model L2.2); 3) malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or reported fever within 48 hours (Model L2.3); and 4) malaria parasitaemia with/without concurrent fever or history of fever (Model L2.4). The rationale for combining CRP with parasitaemia in Model L2.1 was to identify the early-stage of the inflammatory response (incubation or early convalescence phase) using CRP, and late-stage or chronic inflammation (late convalescence phase or asymptomatic malaria) using parasite density (D.I. Thurnham & McCabe, 2012).

All linear models were used to predict iron status in a validation dataset, consisting of one third of the data that were not used in model fitting. The analyses were repeated 100 times, leaving out a different third of the data each time. Descriptive statistics and histograms of the root mean squared errors (RMSE) across analyses for each model were generated. RMSE represents the difference (or standard deviation of the differences) between values predicted by a model and those actually observed. A smaller RMSE would indicate less prediction error or more accurate predictive ability.

3.3.2 Linear modeling and simulation analysis results and conclusions

In the first set of linear models, Hb was significantly associated with ferritin when adjusted for CRP using the covariate method (Models L1.2), though not significant when ferritin was either uncorrected (p=0.534) or corrected for CRP using the regression method (p=0.406) (**Table 3.1**). When simulation analyses were used to compare each linear model before and after including the Hb covariate, the mean of the root mean squared error (RMSE) from Model L1.1 decreased slightly (-0.001) when Hb was added, while the RMSE from Model L1.2 decreased by a greater amount (-0.004), indicating improved predictive ability of the model. Conversely, in Model L1.3, the RMSE increased by 0.01 when Hb was added, indicating that the model may have been over-fitted.

Table 3.1: Comparison of the relationship between baseline ferritin and haemoglobin (Hb) concentrations without (Model L1.1) and with adjustment for inflammation using the covariate method (Model L1.2) or regression method (Model L1.3) among young Ghanaian children (n=1943)

Linear Models	Hb Coefficient (estimate ± SE)	P-value	Model adj R^2	Predictive RMSE (mean ± SE)
Model L1.1				
Without Hb			0.055	1.264 ± 0.003
With Hb	0.015 ± 0.024	0.534	0.055	1.263 ± 0.003
Model L1.2				
Without Hb			0.115	1.223 ± 0.004
With Hb	0.066 ± 0.023	0.005	0.119	1.219 ± 0.003
Model L1.3				
Without Hb			0.055	1.259 ± 0.003
With Hb	0.020 ± 0.024	0.406	0.055	1.269 ± 0.003

Model L1.1: dependent variable was (log) ferritin; covariates were age (<24 or ≥24 months), and sex, with/without baseline Hb

Model L1.2: dependent variable was (log) ferritin; covariates were age (<24 or ≥24 months), sex, baseline CRP (quadratic transformation), with/without baseline Hb

Model L1.3: dependent variable was (log) ferritin adjusted for infection (CRP) using the regression method; covariates were age (<24 or ≥24 months), sex, with/without baseline Hb

CRP = C-reactive protein

SE = Standard Error

Adj R^2 = Adjusted R-squared

RMSE = Root Mean Squared Error (estimated using simulation analyses)

Considering the results from these analyses, it was concluded that compared to ferritin, Hb is not an appropriate biomarker for assessing iron status among children in a malaria endemic area. As demonstrated in Model L1.2, if ferritin is statistically adjusted for the effect of inflammation or infection using the covariate method then the predictive ability of Hb may improve somewhat. Conversely, if ferritin is adjusted for the effect of inflammation or infection using a manual correction method (Model L1.3) then the predictive ability of Hb is diminished. The distinction between Models L1.1 and L1.2 may reflect not only the poor sensitivity of Hb as a an indicator of nutritional iron status, but also the effect of malaria and other infectious diseases (e.g. soil-transmitted helminths, schistosomiasis, HIV/AIDS) on anaemia risk (Balarajan et al., 2011).

In the second set of linear models (**Table 3.2**) infection status was significantly associated with baseline ferritin concentration, whether infection was defined using CRP or parasitaemia or both ($p<0.001$). Similarly, the results from the simulation analyses showed that adding an infection status covariate improved the predictive ability of all models, as the RMSE decreased in all cases (-0.046 for Model L2.1, -0.003 for Model L2.2, -0.017 for Model L2.3, and -0.037 for Model L2.4).

Table 3.2: Comparison of the Relationship between baseline ferritin concentration and infection status, defined as CRP>5 mg/L and/or malaria parasitaemia (Model L2.1); CRP>5 mg/L without malaria parasitaemia (Model L2.2); Malaria parasitaemia with fever (Model L2.3); and Malaria parasitaemia without fever (Model L2.4) among young Ghanaian children (n=1943)

Linear Models	Infection	P-value	Model adj	Predictive

	Coefficient (estimate ± SE)		R^2	RMSE (mean ± SE)
Model L2.1				
Without infection			0.055	1.257 ± 0.004
With infection	0.780 ± 0.060	< 0.001	0.131	1.211 ± 0.003
Model L2.2				
Without infection			0.055	1.257 ± 0.004
With infection	0.420 ± 0.082	< 0 001	0.067	1.254 ± 0.003
Model L2.3				
Without infection			0.055	1.257 ± 0.004
With infection	0.794 ± 0.108	< 0 001	0.080	1.240 ± 0.004
Model L2.4				
Without infection			0.055	1.257 ± 0.004
With infection	0.732 ± 0.070	< 0 001	0.105	1.227 ± 0.003

Model L2.1: dependent variable was (log) ferritin concentration; covariates were age (<24 or ≥24 months), sex, baseline Hb, with/without infection (defined as CRP>5 mg/L and/or malaria parasitaemia)

Model L2.2: dependent variable was (log) ferritin concentration; covariates were age (<24 or ≥24 months), sex, baseline Hb, with/without infection (defined as CRP>5 mg/L without malaria parasitaemia)

Model L2.3: dependent variable was (log) ferritin concentration; covariates were age (<24 or ≥24 months), sex, baseline Hb, with/without infection (defined as malaria parasitaemia with concurrent fever - axillary temperature >37.5^{0}C – or reported fever within 48 hours)

Model L2.4: dependent variable was (log) ferritin concentration; covariates were age (<24 or ≥24 months), sex, baseline Hb, with/without infection (defined as malaria parasitaemia without concurrent fever or history of fever

CRP = C-reactive protein

SE = Standard Error

Adj R^2 = Adjusted R-squared

RMSE = Root Mean Squared Error

These findings highlight the importance of accounting for inflammation or infection when assessing iron status using serum ferritin concentration. Further, assessing infection status using an index variable that combines a biomarker for inflammation (CRP) and parasitaemia (with or without fever) may be more sensitive than using CRP or parasitaemia alone; perhaps due to the fact that CRP is an acute phase protein that represents the early phase of the inflammatory response, while malaria parasitaemia may be present during the later phases, or in some cases without inflammation. Therefore, the combination of infection-related biomarkers could better represent the spectrum of contexts where the assessment of iron status may be confounded.

Overall, the linear regression and simulation analyses demonstrated that accounting for infection may more appropriately improve the predictive ability of statistical models of iron status (defined using plasma ferritin concentration) compared to haemoglobin concentration alone among young children living in rural Ghana. It should be noted that the adjusted R-squared for all linear models tended to be relatively small (0.06-0.13) despite the statistical significance of the associations between Hb, infection and ferritin. This may have been due to the large amount of data used to generate the models (approximately 1900 observations), as larger sample sizes can increase the power for finding significant effects of independent variables with a small amount of explained variance.

3.4 Spatial Model development

The data corresponding to each specific aim were analyzed using generalized linear geostatistical models (GLGM) (P. J. Diggle et al., 1998; P. J. Diggle & Ribeiro, 2006), which were fit using Bayesian inference via an Integrated Nested Laplace Approximation (INLA) algorithm. For geo-spatial analyses, Bayesian methods are preferred over frequentist-based methods, particularly in the case where data are not normally distributed, as the computations required for frequentist inference of non-Gaussian spatial models are virtually intractable (Fong et al., 2010). Considering that the majority of my dependent variables of interest would require non-Gaussian models (e.g. for Aims 2 and 3, as described below in Section 3.4.1), I opted to use Bayesian inference across all analyses in order to maintain consistency in my analytical approach. The decision to use the INLA method was based on its ability to make accurate approximations when applied to *latent* Gaussian models (Rue & Martino, 2009), a subclass of structured additive regression models where the latent field is a Gaussian Markov random field (GMRF) with a sparse precision matrix (Rue & Held, 2005) and a dependent variable that is assumed to be non-Gaussian (Rue & Martino, 2009). Compared to other implementation methods, such as Markov chain Monte Carlo (MCMC), INLA is also more feasible and computationally efficient, especially when it comes to analyzing models with increased complexity (Brooks et al., 2011). This feature of INLA was thought to be beneficial for the analyses described here, which are more exploratory than reductive in nature, with the goal of generating information regarding variable relationships as opposed to achieving parsimony. As such, the models were not restricted in terms of parameterization or complexity, and thus a faster and more computationally convenient implementation method such as INLA was preferred (Rue & Martino, 2009).

Similar reasoning was applied to the choice of the prior distributions. Flat uninformative priors were used for all model coefficients, as this allowed the formation of the posterior distributions to be driven primarily by the data itself. Weak priors were also applied to all model parameters, with the exception of the Matern shape parameter which was fixed at 2. The Matern correlation

function depends on a shape parameter (used to control the differentiability of a continuous spatial surface), as well as a range parameter (controls the rate of change of a spatial surface across a geographical area), and provides a link between Gaussian fields and GMRFs that helps to make non-Gaussian spatial models more computationally efficient (Lindgren, Rue, & Lindstrom, 2011). In the analyses described here, spatial predictions were made on a 100-cell grid covering the study area, and the Matern correlation (approximated by a Markov random field with a buffer of 3 km in each direction) was used to model the spatial random effect. As a result, the residual spatial variation had a normally distributed spatial surface with variance σ^2 and the correlation between the probabilities of the response of two individuals given by a Matern spatial correlation function (where the correlation decreased with distance). .

Given the exploratory nature of the analyses, geo-spatial and non-spatial variables were chosen for inclusion in the final models based on expert opinion and a review of the literature (Aimone et al., 2013), rather than the use of formal model reduction or model selection methods. As illustrated in Figures 1.1 and 1.2, variables were considered for inclusion if they were plausible direct or indirect antecedent factors associated with serum ferritin concentration or infection status (e.g. wealth or elevation). Following this reasoning, and considering the findings from the simulation analyses, haemoglobin concentration was not included as an independent variable; as it was a poor predictor of ferritin concentration, and anaemia is a potential outcome of low iron status or infection.

Selected variables were included in the spatial analyses in five modeling steps. Models 1-3 included all candidate variables grouped by measurement level: Model 1 included individual-level variables only (e.g. age and sex); Model 2 included only household-level variables (e.g. asset score and any distance variables measured from the study compounds); and Model 3 included satellite-derived variables (e.g. elevation and normalized difference vegetation index). These three separate modeling steps were conducted in order to gain additional insight into how

the candidate variables at each level related to the dependent variable. The results from these analyses were also used to inform the development of the final model (Model 4). Variables included in the final model were chosen based on their general order of candidacy, which was based on the strength potential of their relationship with the dependent variable, determined either empirically (e.g. by assessing geographical variation or through exploratory linear regression modelling) or theoretically (e.g. with support from the literature). A 'maximal' model (Model 5) was also developed as a confirmatory modelling step, which included the same covariates as the final model, as well as additional candidate variables selected based on the specific aim of the analysis.

The purpose of the maximal model was to confirm whether the covariates included in the final model explained a sufficient amount of spatial variability in the outcome (e.g. by comparing the spatial random effects); as well as to explore variable relationships that were lower in the order of candidacy, and thus were not included in the final model to preserve statistical power. Although (in practice) it is very difficult to calculate the statistical power of a spatial model, given that spatial data are defined as realizations of a stochastic process (Blangiardo et al., 2013) as opposed to simple point estimates, geostatistical modelling is still in essence a data-fitting process that is vulnerable to sample size limitations and over-parameterization. In light of these risks, the number of covariates (particularly satellite-derived variables) included in the final models were restricted and the remainder of the candidate variables were included in the maximal models. Therefore, the maximal model served as a sort of 'idealistic' model that could also lead to hypothesis-generation, and thus inform future geostatistical analyses.

All spatial models also included a compound-level random effect, as it was expected that the compound-level randomization used in the Ghana trial would likely result in spatial clustering of the outcome at the compound level (albeit a low amount, given the small average cluster size). The results of all spatial models were reported as the exponential of the median estimate of the posterior distribution (e.g. exponential of the median relative risk for the iron status models or odds ratio for most infection models). The median of the posterior distribution was

preferred over the mean, as the data at times showed instability upon exponentiation (i.e. a large skew), resulting in a mean estimate value that fell outside of the 95% credible interval. All spatial modeling was conducted using the *glgm* function from the **geostatsp** package (Brown, 2015) in R (version 3.2.2, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org). The following section includes more specific model descriptions by aim. The corresponding generic model equations can also be found in **Appendix C**.

3.4.1 Model descriptions by aim

For **Aim 1** (Chapter 4), the dependent variable for the baseline and endline spatial models was serum ferritin concentration (log-transformed), corrected for inflammation (CRP) using the regression method (P.S. Suchdev et al., 2016). *Baseline* analyses included 1943 observations, corresponding to the number of compounds from the Ghana trial that were geocoded during the primary data collection phase (section 3.1.3). Selected variables included in Model 1 were baseline age in months (with a change point at 24 months) and sex (male=1, female=2) (**Table 3.3)**. Model 2 consisted of household asset score, maternal education, distance from each study compound to the nearest health facility, and distance from each compound to the District centre. Five satellite-derived variables were included in Model 3: elevation, land cover type (LC), normalized difference vegetation index (NDVI), and two interaction terms for NDVI and the two 'comparator' LC categories (woody savannahs, urban and built up land). The final baseline model (Model 4) combined selected variables from Models 1-3, including age, sex, asset score, distance to the nearest health facility, distance to the District centre, and LC. The 'maximal' model (Model 5) included the same variables as the final model with the addition of maternal education, NDVI, elevation, and the two NDVI-LC interaction terms.

Despite the known up-regulating effect of infection on ferritin concentration, as well as the significant associations observed in the simulation analyses (Table 3.2), infection status was not included as an independent variable in either the baseline or endline ferritin models. The reasons

for this exclusion were two-fold: 1) Because infection was considered to be a potential "outcome" of variations in iron status, either due to depressed immune function (in the case of iron deficiency) or due to potential adverse effects of the iron intervention (in the case of iron sufficiency); and 2) Because ferritin was corrected for inflammation (CRP), which was intended to account for all causes of infection whether malaria or non-malaria related. Furthermore, participants who were diagnosed with malaria at baseline (using an antigen test) were treated before enrolment, which may have introduced bias into the risk relationship with iron status at endline.

Table 3.3: Independent variables included in each baseline (Models 1-5) and endline (Models 4 and 5) spatial model of iron status

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Age (months)	☑			☑	☑
Sex (male reference)	☑			☑	☑
Assets		☑		☑	☑
Maternal education		☑			☑
Distance to health facility (km)		☑		☑	☑
Distance to district centre (km)		☑		☑	☑
Elevation (m)			☑		☑
Urban/built-up land (LC13)			☑	☑	☑
Woody savannahs (LC8)			☑	☑	☑
Normalized difference vegetation index (NDVI)			☑		☑

NDVI*LC8	☑		☑
NDVI*LC13	☑		☑
Group[1]		☑	☑
Baseline iron status[1]		☑	☑
Group*Baseline iron status[1]		☑	☑

[1]Included in endline models (Models 4 and 5) only.

NDVI*LC8 = interaction term between NDVI and LC=8

NDVI*LC13 = interaction term between NDVI and LC=13

The *endline* analyses included 1781 observations, representing trial participants with geocoded compounds who provided blood samples at endline (n=1815) and had both baseline and endline ferritin values (**Figure 3.1**). This amount of attrition (8%) was not considered to be substantial, and statistical differences were not observed when the baseline characteristics of those who were lost to follow up were compared to the full analytic sample at endline (**Appendix D**). For the endline analyses, only the fourth and fifth modelling steps were conducted, and the models included the same variables as the baseline Models 4 and 5 with the addition of baseline iron status (baseline serum ferritin concentration corrected for baseline CRP using the regression method), intervention group (Iron vs. No-iron), and a group-baseline iron status interaction term. Iron status probabilities, after transformation with an identity link function, were modeled as the sum of the contributions of the explanatory variables and a spatially correlated and compound-level random effect terms. A spatially continuous (or geostatistical) model was used for the spatial random effect term, with the correlation between the (log) ferritin concentration of two individuals given by a Matern spatial correlation function applied to the distance separating their respective compounds. Each model was analyzed using the *glgm* function in R (Brown, 2015; R, 2015).

For **Aim 2**, four separate spatial models were created for each *baseline* infection status dependent variable: 1) inflammation (baseline CRP >5 mg/L) with/without baseline malaria parasitaemia; 2) inflammation (baseline CRP >5 mg/L) without baseline parasitaemia; 3) baseline parasitaemia with measured concurrent fever (axillary temperature >37.5^{0}C) or reported history of fever (within 48 hours); and 4) baseline parasitaemia with or without concurrent fever or history of fever. A total of 1943 observations were included in each model, and all dependent variables were binary-valued (coded as '1' for positive infection status). The variables selected for inclusion in the baseline models (**Table 3.4**) included baseline child age in months (with a change point at 24 months), sex, z-scores for weight-for-length and length-for-age, and iron status (baseline serum ferritin corrected for CRP using the regression method, and re-scaled using the inverse of the inter-quartile range) (Model 1). Model 2 included asset score, maternal education, and straight-line (Euclidean) distance from each compound to the nearest health facility. The satellite-derived variables included in Model 3 were elevation, two land cover (LC) types (woody savannahs and urban and built up land), normalized difference vegetation index (NDVI), and two interaction terms for NDVI and each land cover type. The final model (Model 4) included age, sex, weight-for-length z-score, length-for-age z-score, asset score, distance to the nearest health facility, elevation, and iron status. The maximal model (Model 5) included the same variables as the minimal model with the addition of maternal education, NDVI, LC, and the two NDVI-LC interaction terms. . Infection probabilities, after transformation with a logit link function, were modeled as the sum of the contributions of the explanatory variables and a spatially correlated and compound-level random effect terms. A spatially continuous (or geostatistical) model was used for the spatial random effect term, with the correlation between the log-odds of infection of two individuals given by a Matern spatial correlation function applied to the distance separating their respective compounds. Each model was analyzed using the *glgm* function in R (Brown, 2015; R, 2015).

Table 3.4: Independent variables included in each baseline (Models 1-5) and endline (Model 4) spatial model of infection status

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Age	☑			☑	☑
Sex (male reference)	☑			☑	☑
Length-for-age z-score	☑			☑	☑
Weight-for-length z-score	☑			☑	☑
Baseline iron status	☑			☑	☑
Assets		☑		☑	☑
Maternal education		☑			☑
Distance to health facility (km)		☑		☑	☑
Elevation (m)			☑	☑	☑
Urban/built-up land (LC13)			☑		☑
Woody savannahs (LC8)			☑		☑
Normalized difference vegetation index (NDVI)			☑		☑
NDVI*LC8			☑		☑
NDVI*LC13			☑		☑
Baseline infection status[1]				☑	
Group[1]				☑	
Group*Baseline iron status[1]				☑	
Group*Baseline infection status[1]				☑	

[1]Included in endline models (Model 4) only.

NDVI*LC8 = interaction term between NDVI and LC=8

NDVI*LC13 = interaction term between NDVI and LC=13

Similar to Aim 2, four separate spatial models were created for **Aim 3**, this time for 4 definitions of *endline* infection status (i.e. infection status at the end of the intervention period): 1) inflammation (endline CRP >5 mg/L) with/without endline malaria parasitaemia; 2) inflammation (endline CRP >5 mg/L) without endline parasitaemia; 3) endline parasitaemia with measured concurrent fever (axillary temperature >37.5^{0}C) or reported history of fever (within 48 hours); and 4) endline parasitaemia with or without concurrent fever or history of fever. Similar to the baseline infection models, all dependent variables were binary-valued (coded as '1' for positive infection status) with the exception of the third definition (parasitaemia with fever), which was a count variable. The endline datasets used for the infection analyses were similar to those used for the iron status analyses with some variations in sample size due to differences in data collection methods or incongruency between baseline and endline measures. For example, the analytic sample for infection definitions 1 and 2 (using CRP with or without parasitaemia) consisted of 1780 observations, representing trial participants with geocoded compounds who provided blood samples at endline and had both baseline and endline CRP values. For definition 3 (parasitaemia with fever), the 1939 observations included in the Poisson regression analysis corresponded to the number of children with geocoded compounds who had at least one recorded follow-up visit during the intervention period, and thus contributed data to the malaria count outcome.

Unlike the baseline analyses, only the final model (Model 4) was analyzed (**Table 3.4**) with three additional sub-modelling steps: 1) No-iron group only; 2) Iron group only; 3) both groups combined. The interventions groups were analyzed separately in order to differentiate the effect of time and iron treatment on baseline associations observed in Aim 2. The variables included in the by-group models included baseline child age in months (with a change point at 24 months),

sex, baseline z-scores for weight-for-length and length-for-age, asset score, straight-line (Euclidean) distance from each compound to the nearest health facility, elevation, baseline iron status (baseline serum ferritin concentration corrected for baseline CRP using the regression method, and re-scaled using the inverse of the inter-quartile range), and baseline infection status (according to the definition of the dependent variable). The combined-group model included the same variables as the by-group models with the addition of group allocation, and two interaction terms for group-baseline iron status and group-baseline infection status.

Infection probabilities, after transformation with a logit or log link function (for binary or count outcomes, respectively), were modeled as the sum of the contributions of the explanatory variables and spatially correlated and compound-level random effect term. A spatially continuous (or geostatistical) model was used for the spatial random effect term, with the correlation between the log-odds or log-risk of infection of two individuals given by a Matern spatial correlation function applied to the distance separating their respective compounds. Each spatial model was analyzed using the *glgm* function in R (Brown, 2015). In addition to the effect of time and iron treatment, the effect of space was also explored by re-analyzing the combined-group model using a non-spatial generalized linear mixed model with Bayesian inference. The non-spatial model was analyzed using the **INLA** package in R (R, 2015; Rue & Martino, 2009).

3.4.2 Model output plots

Maps of predicted probabilities and residual spatial variation were plotted and overlaid with a base map of the trial area and coordinate points for the compound locations. Predicted probabilities were computed as the posterior means of the dependent variable (after transformation with the applicable link function, see Appendix C), assuming baseline values for individual-level covariates and location-specific values for the spatial covariates. The residual spatial variation plot was the posterior mean of the spatial random effect, corresponding to the

difference between the predicted and expected probability of the outcome at each location (given the spatial covariate at each location). The plots from the final models (Model 4) were visually compared to each other (e.g. between baseline and endline or across dependent variable definitions) and to relevant satellite-derived maps (e.g. elevation), where applicable, in order to generate potential explanations for the spatial patterns observed.

Chapter 4
Aim 1 Manuscript

[Note: To be formatted for submission to Journal of Nutrition: http://jn.nutrition.org/site/misc/instructions-for-authors.xhtml]

4 Geo-spatial patterns of iron status among young children in rural Ghana before and after participating in a randomized trial of iron home-fortification

Ashley M. Aimone Phillips, Patrick E. Brown, Seth Owusu-Agyei, Donald

C. Cole, Stanley H. Zlotkin

Abstract

Background: There is a need to identify biomarkers or indicators of iron deficiency risk that are not confounded by infection or inflammation, and that are feasible to measure at a population level. We conducted a secondary spatial analysis of baseline data from a cluster-randomized trial (conducted in 2010) investigating the impact of iron home-fortification on malaria risk in young Ghanaian children to explore the utility of model-based geostatistics as an alternative means of assessing iron status.

Objective: To determine the geo-spatial factors associated with the variation in ferritin concentration among young Ghanaian children before and after participating in a 5-month iron intervention trial.

Methods: Spatial analyses were conducted using a linear geostatistical model with a Matern spatial correlation function. Potentially informative variables were included in the final baseline

and endline models through five modelling steps: individual-level variables (Model 1); household-level variables (Model 2); satellite-derived spatial variables (Model 3); and a minimal and maximal set of covariates from Models 1 to 3 (Models 4 and 5, respectively).

Results: At baseline, in Models 4 and 5, iron status was significantly associated with age, sex, asset score, and maternal education. At the end of the intervention period, iron status was associated with age below 24 months, sex, and baseline iron status. None of the environmental factors included in the models demonstrated significant associations with iron status at baseline or endline; however the large spatial random effects indicated significant spatial variation in iron status across the study area.

Conclusions: The geographical variables included in our models did not explain a significant amount of variation in iron status after controlling for age, sex, wealth, and maternal education. The large residual spatial variation indicates that location may play a role in iron deficiency risk; however, it is difficult to quantify the potential confounding effect of infection.

Introduction

Iron is an essential nutrient that all children require for physiological growth and development (1). Inadequate intake of dietary iron can lead to iron deficiency, which is estimated to account for 20,854 deaths and 2.2 million disability-adjusted life years lost per year among children less than five years of age, most of whom live in low- and middle-income countries (LMICs) (2). Iron deficiency is associated with adverse developmental outcomes such as impairments in cognitive function, motor development (3), growth (4), and immune function (1). When iron deficiency is left untreated, it can lead to anaemia, which is characterized by low blood haemoglobin concentration. Approximately 42% of all anaemia cases are attributed to nutritional iron insufficiency (5), and thus haemoglobin (Hb) concentration has commonly been used to estimate the prevalence of iron deficiency at the population level (6, 7).

A major limitation of using Hb to assess iron status is that it is not a very specific or sensitive biomarker of iron homeostasis. For example, it takes a loss of approximately 20-30% of iron stores before Hb levels fall below the "anaemia" threshold (8), thus resulting in a potentially large undetected population of iron deficient non-anaemic children. Haemoglobin can also decrease as a result of infectious diseases such as malaria, genetic causes (e.g. sickle cell or thalassemia), or other micronutrients deficiencies (e.g. vitamin B12, folate, or vitamin A) (9). More specific biomarkers of iron status such as serum ferritin (a measure of body iron stores) are available (10); however, they are less practical and feasible to measure at a population level compared to Hb. Furthermore, inflammation (due to infection) has an up-regulating effect on ferritin, an acute phase protein (11), and acute or chronic inflammation may have an adverse effect on iron homeostasis (12) or the risk of iron deficiency (13). To add to this complexity, evidence from a large randomized trial conducted in Pemba, Zanzibar in 2003 indicated that in malaria endemic areas, supplementing young children with iron may increase their risk of malaria and infection-related morbidity and mortality, particularly if they are iron replete (14). Considering that recent estimates of malaria prevalence have suggested that it accounts for 584,000 deaths worldwide (most of which occur in Africa and among children less than 5 years of age) (15), the findings from Pemba highlight the significance of the "iron and malaria"

problem, as well as the need to identify biomarkers or indicators of iron deficiency risk that are not confounded by infection or inflammation, and that are feasible to measure at a population level.

Geographical factors, such as the environmental or spatial characteristics of a community or region, may provide additional insight into the dynamics and distribution of iron status among targeted populations, such as children in LMICs. Knowledge of the geographical patterns and spatial factors associated with iron deficiency could inform treatment needs (16, 17), and thus help implementers and policy makers in LMICs to determine where health services should be targeted in order to make efficient use of limited resources. Collecting geo-spatial data is also non-invasive and less costly compared to biological indicators; and are often publicly available, which improves the access to and comparability of population-level trends across regional and national borders. Geographical information systems (GIS) and spatial analysis methods have been used in several areas of health research, such as disease surveillance and risk analysis, assessment of health system access, and health system planning (18-21). In terms of nutrition research, however, the investigation of geo-spatial risk factors associated with iron deficiency among children in LMICs has been limited (22), and almost entirely restricted to the use of haemoglobin concentration or anaemia status as a primary outcome (22-25). Considering the risks of providing iron supplements to children in malaria endemic areas who may be anaemic (e.g. due to malaria or other infections) though not iron deficient (due to poor dietary intake) (14), it is important to be able to differentiate these states in order for the benefits of iron deficiency control programs to outweigh their risks and costs.

The current state of evidence does not sufficiently support our understanding of how spatial analyses can be utilized to identify populations at risk of (true) nutritional iron deficiency in LMIC settings. In order to fill this gap in the literature, the primary objective of the analyses reported here was to determine the geo-spatial factors associated with the variation in ferritin concentration among young children before and after participating in a 5-month randomized iron intervention trial in rural Ghana.

Methods

Study population

The data used in these analyses were generated from a community-based cluster randomized trial conducted in 2010 in a rural area of Ghana (Brong-Ahafo Region) during the rainy season (March to November) (26). At the time, Ghana had an estimated 7.2 million cases of malaria per year, and the prevalence of anaemia among preschool aged children was 76.1% (95% CI 73.9-78.2%) (7, 27). More recent estimates indicate that the incidence of malaria decreased between 2006 and 2013 (1.6 million cases in 2013), though the country is still considered to be a high transmission area (>1 case per 1000 population) (15). Conversely, the prevalence of anaemia in Ghana remained stable between 2003 and 2011 (76.0%, 95% CI 66.0-83.0% in 2011) and the public health significance continues to be "severe" (5, 7). Briefly, the aim of the clinical trial was to determine the effect of providing iron in powder form with other micronutrients (micronutrient powders) for 5 months during the rainy season on the incidence of malaria among 1958 children aged 6-35 months (representing 1552 clusters and 22 villages) (26). Randomization was conducted at the level of the compound, which was defined as a residence comprising of one or more households. The study area covered approximately 3,800 km^2, and is depicted in **Figure 4.1**, including study compounds, health facilities, and road networks.

Measures from trial data

Biological samples collected at baseline and endline were analyzed for biomarkers of iron and inflammation or infection status, including serum ferritin (SF), C-reactive protein (CRP), and malaria parasite density. Plasma CRP and plasma ferritin were measured using an immunoturbidimetric method (QuickRead CRP, Orion Diagnostica, Espoo, Finland) and an enzyme immunoassay (Spectro Ferritin S-22, Ramco Laboratories Inc., Stafford, USA), respectively. Distributions of serum ferritin were positively skewed ($D=0.24$, $p<0.01$ for the

Kolmogorov-Smirnov test of normality), and thus log transformed values were used in all analyses [see Zlotkin et al. (2013) for a complete list of biochemical measures]. Demographic and nutrition-related information was collected at baseline at the household and individual levels, including household assets, maternal education, and child feeding practices.

Geographical coordinates

Approximately 24 months after completion of the clinical trial (August-November 2012), geographical coordinates for over 95% of the study compounds (representing 1943 trial participants), 22 study villages, and surrounding health facilities and road networks were collected using handheld global positioning system (GPS) units. The GPS coordinates were measured using the WGS 1984 coordinate system and transformed to a universal transverse Mercator (UTM) Zone 30N projection (EPSG code: 32630).

Satellite-derived data

Elevation data (SRTM DEM Version 3) were downloaded from the U.S. Geological Survey (USGS) (28), which is a government agency for the natural sciences that provides information on the global environment, such as the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) data sets. The data file had a resolution of 3 arc-seconds (approximately 90 meters) and covered approximately 80% of Earth's total land mass (between 56 degrees south and 60 degrees north latitude). Normalized Difference Vegetation Index (NDVI) data were downloaded from the Land Processes Distributed Active Archive Center (LPDAAC) (29). The NDVI data was produced by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument (product MOD13Q1), which uses blue, red, and near-infrared reflectances to determine vegetation indices for 16-day intervals with a 250-meter spatial resolution. The MODIS land cover type product (MOD12Q1) was downloaded from worldgrids.org and consisted of 17 land cover classes (defined by the International Geosphere Biosphere Programme or IGBP), including natural vegetation (11 classes), developed and mosaic land (3 classes), and non-vegetation (3 classes). Land cover data were derived from

yearly output from the Terra and Aqua MODIS instruments at a resolution of 500 meters (30).

Spatial modelling

The data were geocoded at the compound-level and analyzed using linear geostatistical models (LGM) (31, 32). The dependent variable for the spatial models was log-transformed serum ferritin concentration. In order to account for the effect of malaria and non-malaria infection, ferritin values were corrected for inflammation (baseline CRP) using the regression method (33). Geo-spatial and non-spatial variables were chosen for inclusion in the final models based on expert opinion and a review of the literature (22). Variables were considered for inclusion if they were plausible direct or indirect antecedent factors associated with iron deficiency (e.g. wealth status), and excluded if they were outcomes of iron deficiency (e.g. anaemia). Iron status probabilities, after transformation with an identity link function, were modeled as the sum of the contributions of the explanatory variables and a spatially correlated and compound-level random effect terms. A spatially continuous (or geostatistical) model was used for the spatial random effect term, with the correlation between the probability of higher ferritin concentrations of two individuals given by a Matern spatial correlation function applied to the distance separating their respective compounds.

The models were fit using Bayesian inference via an Integrated Nested Laplace Approximation (INLA) algorithm (34). Weak or uninformative priors were used for all model parameters with the exception of the Matern shape parameter, which was fixed at 2. Spatial predictions were made on a 100-cell grid covering the study area, and the Matern correlation approximated by a Markov random field (35) extended an additional 3 km in each direction.

Selected variables were included in the baseline spatial analyses in five modelling steps. Model 1 included only the individual-level variables enrolment age and sex. Age in months was calculated using the reported date of birth and trial enrolment date, and was modeled as a continuous linear effect with a change point at 24 months as this was the closest half-year to the

mean age of those children who were no longer receiving breast milk (mean=26.8 months ± 5.8, n=746). A change point of 24 months has also been used in other studies of iron deficiency and anaemia in children (36, 37). Model 2 included only household-level variables: asset score, maternal education, distance from each study compound to the nearest health facility, and distance from each compound to the District centre (38). Household asset score was generated using a principal component analysis of 6 economic indicators, including farm ownership, size and type of crops grown (e.g. cash crop for sale in the market or for household consumption), type of toilet facility, and house ownership (e.g. own vs. rent). For descriptive purposes, asset score was dichotomized at the median. Maternal education was included as a dichotomous variable for "none" versus "any" education. Distance to the nearest health facility was included as an indicator of access to the health care system, a factor that may influence iron status through the distribution and use of iron supplements or the treatment of other illnesses that can affect iron intake or utilization (e.g. reduced appetite due to fever or reduced absorption due to infection). Distance to the District centre (Wenchi Town) was included as an indicator of access to other community-level factors that can affect health (e.g. local markets, schools, pharmacies, and employment opportunities) (38). All distances were measured `as the bird files' (or Euclidean) using the Near Table tool in ArcMap (ArcGIS 10.2, Environmental Systems Resource Institute, Redlands, California).

Model 3 included five satellite-derived variables: elevation, land cover type (LC), normalized difference vegetation index (NDVI), and two interaction terms for NDVI and LC. All satellite-derived data were downloaded as global datasets and cropped according to the geographical boundaries of the study area. Elevation had a range of 116-530 meters, and values were centered by subtracting 250 before including the raster in the analyses. Land cover type was a discrete categorical variable consisting of 3 values representing woody savannahs (LC=8, n=21/1943 observations), urban and built up land (LC=13, n=243/1943 observations), and cropland/natural vegetation mosaic (LC=14, n=1679/1943 observations). For modelling purposes, the largest category (cropland/natural vegetation mosaic) was used as the reference. NDVI was included as a proxy for soil moisture (39), and was averaged over the year that the study was conducted (2010) in a single raster file. Averaged NDVI values had a range of 0.22-0.62. Two NDVI-LC

interaction terms were created by centering the average NDVI raster values (NDVI / 1000 – 4), and adjusting the resolution of the LC raster to match that of the NDVI raster. The NDVI raster was then used to mask the LC raster except where LC=8 (these cells were given a value of zero). This created a new raster that could be used to investigate whether the association between the dependent variable and vegetation (or soil moisture) varied across areas with or without a woody savannah land cover type. The same method was applied for the NDVI-LC interaction term with LC values of 13 (urban and built-up land).

The final baseline model (Model 4) combined selected variables from Models 1-3, including age, sex, asset score, distance to the nearest health facility, distance to the District centre, and LC. A 'maximal' model (Model 5) was also developed as a confirmatory modelling step and included the same variables as the final model with the addition of maternal education, NDVI, elevation, and the two NDVI-LC interaction terms. Similar to the baseline models, the dependent variable for the post-intervention analyses was endline serum ferritin concentration (log transformed and corrected for endline CRP using the regression method). Only the fourth and fifth modelling steps were conducted (Models L4 and L5), and the models included the same variables as the baseline Models 4 and 5 with the addition of baseline iron status (baseline serum ferritin concentration corrected for baseline CRP using the regression method), intervention group (Iron vs. No-iron), and a group-baseline iron status interaction term. All spatial modelling was conducted using the *glgm* function from the "geostatsp" package in R (40, 41).

Maps of predicted iron status probabilities and residual spatial variation from the final baseline and endline models were plotted and overlaid with a base map of the trial area and coordinate points for the compound locations. Predicted probabilities were computed as the posterior means of ferritin concentration, assuming baseline values for individual-level covariates and location-specific values for the spatial covariates. The residual spatial variation plot represents

the posterior mean of the spatial random effect, corresponding to the difference between the predicted and expected ferritin concentration at each location (given the spatial covariate at each location).

Ethics

Approval for the original clinical trial was obtained from the Kintampo Health Research Centre (KHRC) Institutional Ethics Committee, the Ghana Health Service (GHS) Ethical Review Committee, the Hospital for Sick Children Research Ethics Board, and the Food and Drugs Board of Ghana. Approval for conducting the secondary analysis of trial data, as well as the collection and primary analysis of geographical data was obtained from the Hospital for Sick Children and University of Toronto Health Sciences Research Ethics Boards.

Results

Baseline and endline characteristics of the study sample are presented in **Table 4.1**, including biochemical and anthropometric measures, and demographics. A total of 1943 trial participants with geocoded compounds were included in the baseline analyses. At endline, the analytical sample consisted of 1781 observations, representing trial participants who provided a blood sample at endline (n=1815), had a geocoded compound, and had both baseline and endline ferritin values. This amount of attrition (8%) was not considered to be substantial, and statistical differences were not observed when the baseline characteristics of those who were lost to follow up (n=162) were compared to the full analytic sample at endline. The mean age at enrolment was 19.2 months, with 69% (1348/1943) of participants aged below 24 months. After correcting ferritin for the effect of inflammation (CRP) using the regression method, the prevalence of iron deficiency (ferritin <12 µg/L) was 21.4% (415/1943) at baseline and 9.21% (164/1781) at endline (70/886 in the Iron group, 94/895 in the No-iron group). Contrary to iron deficiency, the proportion of children with acute inflammation (CRP>5 mg/L) increased slightly from baseline

(25.3%, 491/1943) to endline (29.1%, 518/1781), particularly in the Iron group (270/1781 versus 248/1781 in the Iron and No-iron groups, respectively). In terms of parasitaemia, 23.0% (447/1943) of trial participants had a blood parasite density >0/µL (with or without fever) at baseline, while 27.1% (483/1781) had parasitaemia at endline (13.9% in the Iron group and 13.2% in the No-iron group).

The results of the baseline spatial models have been summarized in **Table 4.2**. In Model 1 (individual-level) age and sex were positively associated with ferritin concentration, meaning that younger male children tended to have poorer iron status at baseline compared to their older or female counterparts. In Model 2 (household-level), asset score was positively associated with ferritin concentration, while maternal education had an inverse association. These results suggested that children from wealthier households had approximately 8% higher ferritin concentration than those of poorer families [relative risk (RR) 1.08, 95% credible interval (CrI) 1.01, 1.16]. On the contrary, mothers with a higher education tended to have children with 14% lower ferritin concentrations compared to those with less education (RR 0.86, 95% CrI 0.75, 0.98). None of the variables included in the spatial model (Model 3) were significantly associated with baseline ferritin concentration. In the final model (Model 4), age, sex, asset score remained significantly (and positively) associated with ferritin concentration. When maternal education, elevation, NDVI, and the NDVI*LC interaction terms were added to the covariate list (Model 5), distance to the nearest health facility became significant (along with age and sex), and the relationship with ferritin was positive, meaning that every kilometer of distance from a health facility corresponded with a 7% higher ferritin concentration (RR 1.07, 95% CrI 1.00, 1.14).

The results of the endline spatial models have been summarized in **Table 4.3**. In Model L4 iron status at the end of the intervention period was positively associated with age (6-23 months) and sex. These findings indicate that female children tended to have 12% higher ferritin concentration than their male counterparts, and older children between 6 and 23 months were more at risk of iron deficiency. Baseline iron status was also positively associated with iron

status at endline. This suggests that every one unit increase in serum ferritin concentration at enrolment likely translated into a 15% higher ferritin measurement at endline. Model L5 showed similar results to Model L4, with the exception of sex, which was still positively associated with endline iron status though no longer statistically significant (RR 1.11, 95% CrI 0.99, 1.25).

All models demonstrated significant residual spatial variation in iron status at baseline and endline across the study area. On the contrary, the compound random effects tended to be small with narrow 95% credible intervals, indicating low variability in iron status between compounds. At baseline, the distance at which the inter-variable relationships started to decay (decreased covariance) ranged from 3.61 (95% CrI 1.84, 6.79) kilometers in Model 1 to 4.76 (95% CrI 2.16, 9.38) kilometers in Model 3. At endline, the range of both Models L4 and L5 was approximately 4 kilometers. Plots of the predicted mean and residual spatial variation from Models 4 and L4 are presented in **Figures 4.2a/b** and **4.3a/b**, respectively. Both figures show defined high- and low-risk areas that seem to cluster around villages. At endline, the high-risk clusters appear to be larger, particularly around the District centre. The plots depicting the spatial random effect from each model (**Figures 4.2b and 4.3b**) are similar to those of the predicted posterior mean (**Figures 4.2a and 4.3a**), indicating that the predicted spatial relationships may have been primarily driven by residual variation (i.e. spatial variation that was not fully explained by the variables included in the models).

Discussion

The analyses conducted herein explored the spatial variation and associated risk factors of iron status among trial participants before and after a 5-month iron intervention. At baseline, age and sex were consistently and positively associated with ferritin concentration, whereas at endline age (at enrolment) was only a significant risk factor for those between 6 and 23 months of age.

The association with sex was also weakened at endline when maternal education, elevation and NDVI were added to the model (Model L5). Since higher ferritin values indicate increased body iron stores, these results suggest that younger male children were more likely to be iron deficient at baseline, and younger children who were less than 24 months old at enrolment were more likely to have lower iron stores at endline.

Younger children may have had poorer iron status due to prolonged breastfeeding or inadequate intakes of dietary iron from weaning foods. Although the reported age of introduction of complementary foods was similar across age groups, through qualitative data collection we found that mothers and caregivers typically added the micronutrient powders to corn- or rice-based foods (e.g. *koko, kenkey, banku, tombrown, rice porridge*) (data not shown), which tend to have low iron content and high amounts of plant-derived compounds, such as phytates, that reduce the bioavailability of iron (42, 43). These findings are supported by other population-level surveys, such as the 2008 Ghanaian Demographic and Health Survey (DHS), which indicated that the most common types of complementary foods given to children under 24 months of age who were still breastfeeding included grains (85%), roots and tubers (40.7%), and other fruits and vegetables (52.9%) (44). Furthermore, results from the DHS showed that the proportion of breastfeeding children who received meat, fish, poultry, and eggs increased from 64.6% to 87% among those who were 6-23 and 24-35 months of age, respectively (44). Therefore, older children were more likely to have higher intakes of bioavailable iron, a trend that also supports the higher probability of low ferritin concentrations among children less than 24 months of age observed in the current analysis.

Greater iron deficiency and anaemia risk among male children compared to females has also been reported in other studies from various African countries (23, 45-49), as well as a review by Kassebaum et al. (2014) on the global anaemia burden between 1990 to 2010 across 187 countries (50). Various explanations for this relationship have been suggested, including hormonal or genetic factors that result in a higher susceptibility to anaemia among males compared to females (50), lower iron stores at birth or higher intestinal losses among male

children (51), as well as misclassification (46). Among these potential explanations, however, we are not aware of any that have been specifically investigated or confirmed.

In the baseline model that included only household-level variables (Model 2), asset score was found to be positively associated with ferritin concentration, meaning that children from wealthier households were more likely to be iron sufficient. This relationship agrees with the well-documented finding that "wealth" is an important underlying factor of maternal and child health in low- and middle-income settings (52). Perhaps more surprisingly, maternal education was negatively associated with baseline ferritin, suggesting that children with more educated mothers were more likely to have poorer iron status. This finding is counter-intuitive because higher maternal education is often associated with better feeding and care practices, and thus healthier children with fewer infections or nutritional deficiencies (53). Following this reasoning, a possible explanation for the inverse relationship observed here pertains to the up-regulating effect of inflammation on serum ferritin concentration. Children with more educated mothers may have indeed had a lower risk of infection, which then reduced the probability that ferritin concentration was falsely elevated and potentially "unmasked" a true iron deficiency that better represented the poor iron content of typical weaning foods (as described above). An exploratory analysis showed that mean ferritin concentration at baseline decreased significantly with increasing maternal education (Pearson R = -0.08, $p<0.001$; p-value for F-test = 0.005); however, this relationship was no longer significant when observations with high CRP (n=491) or outlying ferritin values (n=1) were removed (Pearson R = -0.06, $p>0.05$; p-value for F-test = 0.088).

It should also be noted that the inverse relationship with maternal education was not statistically significant in Model 5 (RR 0.898, CrI 0.79, 1.02), which suggests that the significant association observed in Model 2 may have been due to a Type I error (false positive or false rejection of the null hypothesis). The same logic could be applied to the significant association with distance to the nearest health facility that was observed in Model 5, especially given that the lower bound of the credible interval was very close to 1.00.

In the third baseline model (Model 3), a lack of significant associations may have suggested that a certain amount of spatial variation was not explained by the variables chosen for inclusion in the model. In order to optimize power given our sample size, only three satellite-derived variables were included in the model; however, they were carefully selected from a larger list of potential covariates generated with support from the literature and expert opinion. A larger sample size and/or geographical coverage may have allowed us to include other covariates from this list or strengthened the associations within the existing model. For example, Magalhaes et al. modeled haemoglobin concentration among 24,277 children less than 5 years of age in western Africa and found significant associations with elevation, land surface temperature, normalized difference vegetation index, and distance to a perennial body of water (24). In the current analysis, aside from age, sex, asset score and maternal education, no other variables included in Models 4 and 5 were significantly associated with baseline serum ferritin concentration. Similarly, in the endline models, only a small group of individual-level variables (age, sex, baseline iron status) were associated with endline ferritin concentration.

Although several associations found at baseline were not observed at the end of the intervention period, the spatial variation in iron status remained significant. This suggests that the geographical distribution of iron status may in fact be an important feature of iron deficiency risk among children, even after receiving an iron intervention, though it was not fully explained by the factors included in the endline analyses. It is possible that a larger sample size and/or geographical coverage may have allowed us to gain additional insight into potential sources of spatial variation; however, that this could also have the undesired effect of increase the risk of a Type I error.

In the current analyses, the model range remained relatively consistent at 4 kilometers from baseline to endline, which may represent the size of geographical clusters (24). Although the cluster size did not change appreciably from baseline to endline, the predicted mean plots of

endline iron status demonstrated more "low-ferritin" (or "high-deficiency") risk areas, particularly around the District centre (**Figures 4.2 and 4.3**). This finding is not necessarily expected, since others have found that living in or near a municipal centre is associated with improved health due to greater access to nutrition and health care (54). For example, Shirayama et al. used GIS to create maps of malaria prevalence among 1,711 villagers in Khammouane province, Laos and found a lower prevalence of active cases among villages with greater access to the district centre (38). In the Ghana trial, iron status was assessed using serum ferritin, which is an acute phase protein. Since we were only able to correct ferritin for elevated levels of CRP, it is possible that residual confounding remained. Therefore, if children living near the District centre had a lower risk of infection, their serum ferritin values may have been lower compared to those in more distal villages with potentially higher infection risk. In terms of the effect of the intervention, it is difficult to separate this from the influence of other potential contributing factors. For example, while roughly half of the participants received iron, both groups received additional micronutrients (zinc, vitamin A, vitamin C) that may have led to overall improvements in immune function. Further, all participants were monitored closely throughout the trial and provided with malaria treatment when indicated, as well as improved access to the health system.

Relative to the residual spatial variation, the compound random effects observed in all baseline and endline models suggested low variability in the outcome (baseline or endline iron status) between compounds. In the original trial, the iron treatment was randomised at the compound level, meaning that groups of children in adjacent compounds may have been in separate treatment arms. It was expected that the compound-level randomization would have resulted in spatial clustering of the outcome. There are two potential explanations for why this was not observed: 1) the compounds within a village were in close proximity to each other, which may have lead to clustering of the outcome at the village level despite differences in treatment between compounds; and/or 2) the small number of observations per compound (average cluster size was 1.3) likely reduced the opportunity for the outcome to cluster within compounds, and thus reduced the variability across compounds.

A potential limitation of the present analyses was the incomplete correction of serum ferritin for inflammation or infection. Both ferritin and C-reactive protein (CRP) rise in accordance with the early phase of the inflammatory response (approximately 0-2 days). When CRP declines, however, ferritin remains elevated and more closely approximates other acute phase proteins, such as alpha-1-acid glycoprotein (AGP), that reach their peak concentration during the late phase (approximately 2-8 days) (11). Since only CRP was available for the current analysis, and the study protocol was not designed to include diagnoses of non-malarial infections, it is possible that the prevalence of infection among participants was underestimated resulting in the overestimation of ferritin concentration at baseline and endline.

An additional limitation was the use of Euclidean distance rather than network (travel) distance when estimating proximity to a health facility or the district centre. Network distance may have been a more appropriate indicator of travel impedance or access since it can more appropriately account for travel distance by vehicle or bicycle. Calculating distance by road from all study compounds to the nearest health facility or District Centre was not possible due to incomplete or missing vector information (e.g. miss-aligned junctions, missing or disconnected road segments). We considered using the assumption that all travel distances started at the nearest road, though this was not appropriate for those villages in more remote locations with highly dispersed compounds. Nesbitt et al. (2014) encountered similar challenges in a study comparing different measures of travel impedance to estimate access to delivery care in the Brong-Ahafo region of Ghana (55). After addressing each challenge and creating a detailed network map layer, the authors found that straight-line distance was as informative as network distance for determining spatial access in this setting of rural Ghana (55). Therefore, we concluded that using Euclidean rather than network distance in the present analysis could be justified.

Conclusions

In the current analyses, spatial modelling was explored as an alternative means of assessing the risk of iron deficiency among children in a malaria endemic area that is less invasive and resource-intensive than collecting biological samples. Given the findings, it appears that the spatial variables chosen could not provide additional insights, likely due to the powerful confounding effect of inflammation and infection. This provides further evidence that assessing iron status in children can be challenging and complex, particularly in areas where the risk of malaria and other infections is high. Considering recent concerns with respect to the interaction between iron and malaria, future research should consider spatial analyses of infection status among children in high malaria transmission areas as a means to identify geographic locations where the safety and effectiveness of iron supplementation may be compromised.

References

1. Beard JL. Iron biology in immune function, muscle metabolism and neuronal functioning. J Nutr. 2001;131(2S-2):568S-79S; discussion 80S.
2. Stoltzfus RJ. Iron deficiency: global prevalence and consequences. Food Nutr Bull. 2003;24(4 Suppl):S99-103.
3. Grantham-McGregor S, Ani C. A review of studies on the effect of iron deficiency on cognitive development in children. J Nutr. 2001;131(2S-2):649S-66S; discussion 66S-68S.
4. Lawless JW, Latham MC, Stephenson LS, Kinoti SN, Pertet AM. Iron supplementation improves appetite and growth in anemic Kenyan primary school children. J Nutr. 1994;124(5):645-54.
5. WHO. The global prevalence of anaemia in 2011. Geneva: World Health Organization, 2015.
6. WHO/CDC. Assessing the Iron Status of Populations: report of a joint WHO/CDC technical consulation on the assessment of iron status at the population level. Geneva: World Health Organization; 2005.
7. WHO. Worldwide prevalence of anaemia 1993-2005. Geneva, Switzerland: World Health Organization, 2008.
8. Zimmermann MB. Methods to assess iron and iodine status. Br J Nutr. 2008;99 Suppl 3:S2-9.
9. Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV. Anaemia in low-income and middle-income countries. Lancet. 2011;378(9809):2123-35.
10. Worwood M. Indicators of the iron status of populations: ferritin. In: WHO/CDC, editor. Assessing the iron status of populations: report of a joint World Health Organization/Centers for Disease Control and Prevention technical consultation on the assessment of iron status at the population level. 2nd ed. Geneva: World Health Organization; 2007.
11. Thurnham DI, McCabe GP. Influence of infection and inflammation on biomarkers of nutritional status with an emphasis on vitamin A and iron. In: Organization WH, editor.

Report: Priorities in the assessment of vitamin A and iron status in populations, Panama City, Panama, 15-17 September 2010. Geneva: World Health Organization; 2012. p. 63-80.

12. Spottiswoode N, Duffy PE, Drakesmith H. Iron, anemia and hepcidin in malaria. Front Pharmacol. 2014;5:125.
13. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science. 2004;306(5704):2090-3.
14. Sazawal Z, Black, RE., Ramsan, M., Chwaya, HM., Stoltzfus, RJ., Dutta, A ., et al. Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: community-based, randomised, placebo-controlled trial. Lancet. 2006;367:133-43.
15. WHO. World Malaria Report 2014. Geneva, Switzerland: World Health Organization, 2014.
16. Schur N, Vounatsou P, Utzinger J. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. PLoS Negl Trop Dis. 2012;6(9):e1773.
17. Pullan RL, Gething PW, Smith JL, Mwandawiro CS, Sturrock HJ, Gitonga CW, et al. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. PLoS Negl Trop Dis. 2011;5(2):e958.
18. Nykiforuk C, Flaman L. Exploring the utilization of geographic information systems in health promotion and public health. Edmonton, AB, Canada: School of Public Health, University of Alberta, 2008.
19. Song P, Zhu Y, Mao X, Li Q, An L. Assessing spatial accessibility to maternity units in Shenzhen, China. PLoS One. 2013;8(7):e70227.
20. Masters SH, Burstein R, Amofah G, Abaogye P, Kumar S, Hanlon M. Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis. Soc Sci Med. 2013;93:147-54.
21. Root ED, Lucero M, Nohynek H, Anthamatten P, Thomas DS, Tallo V, et al. Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children. Proc Natl Acad Sci U S A. 2014;111(9):3520-5.

22. Aimone AM, Perumal N, Cole DC. A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: the case of anaemia and malaria. International Journal of Health Geographics. 2013;12:1-13.
23. Magalhães RJ, Clements AC. Mapping the risk of anaemia in preschool-age children: the contribution of malnutrition, malaria, and helminth infections in West Africa. PLoS Med. 2011;8(6):e1000438.
24. Soares Magalhães RJ, Clements AC. Spatial heterogeneity of haemoglobin concentration in preschool-age children in sub-Saharan Africa. Bull World Health Organ. 2011;89(6):459-68.
25. Pullan RL, Gitonga C, Mwandawiro C, Snow RW, Brooker SJ. Estimating the relative contribution of parasitic infections and nutrition for anaemia among school-aged children in Kenya: a subnational geostatistical analysis. BMJ Open. 2013;3(2).
26. Zlotkin S, Newton S, Aimone AM, Azindow I, Amenga-Etego S, Tchum K, et al. Effect of iron fortification on malaria incidence in infants and young children in Ghana: a randomized trial. JAMA. 2013;310(9):938-47.
27. WHO. World Malaria Report 2009. Geneva, Switzerland: World Health Organization, 2009.
28. USGS. United States Geological Survey 2012 [Available from: https://lta.cr.usgs.gov.
29. USGS. Land Processes Distributed Active Archive Center 2014 [Available from: https://lpdaac.usgs.gov.
30. Hengl T. Worldgrids — a public repository and a WPS for global environmental layers: ISRIC - World Soil Information; 2014 [Available from: http://worldgrids.org.
31. Diggle PJ, Moyeed RA, Tawn JA. Model-based geostatistics. Applied Statistics. 1998;47:299-350.
32. Diggle PJ, Ribeiro PJ. Model-based geostatistics. New York: Springer-Verlag; 2006.
33. Namaste S, Rohner R, Suchdev P, Kupka R, Mei Z, Bhushan N, et al. Approaches to address the association between inflammation and ferritin in children and women. Unpublished.

34. Rue H, Martino S. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B. 2009;71(2):319-92.
35. Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society (Series B, Statistical Methodology). 2011;73, Part 4:423-98.
36. Adish AA, Esrey SA, Gyorkos TW, Johns T. Risk factors for iron deficiency anaemia in preschool children in northern Ethiopia. Public Health Nutrition. 1999;2(3):243-52.
37. Jonker FA, Calis JC, van Hensbroek MB, Phiri K, Geskus RB, Brabin BJ, et al. Iron status predicts malaria risk in Malawian preschool children. PLoS One. 2012;7(8):e42670.
38. Shirayama Y, Phompida S, Shibuya K. Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. Malar J. 2009;8:217.
39. Samadoulougou S, Maheu-Giroux M, Kirakoya-Samadoulougou F, De Keukeleire M, Castro MC, Robert A. Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. Parasit Vectors. 2014;7:350.
40. Brown PE. Model-based geostatistics the easy way. Journal of Statistical Software. 2015;63(12):1-24.
41. R. R: A Language and Environment for Statistical Computing. 3.2.2 ed. Vienna, Austria: R Foundation for Statistical Computing; 2015.
42. Raiten D, Namaste S, Brabin B. Considerations for the Safe and Effective Use of Iron Interventions in Areas of Malaria Burden: Full Technical Report. Bethesda, MD: NICHD, NIH, HHS, 2009.
43. Onofiok NON, D.O. Weaning foods in West Africa: Nutritional problems and possible solutions. Food and Nutrition Bulletin. 1998;19(1).
44. GSS/GHS. Ghana Demographic and Health Survey 2008. Accra, Ghana: Ghana Statistical Service (GSS), Ghana Health Service (GHS), ICF Macro, 2009.
45. Yang Z, Lonnerdal B, Adu-Afarwuah S, Brown KH, Chaparro CM, Cohen RJ, et al. Prevalence and predictors of iron deficiency in fully breastfed infants at 6 mo of age: comparison of data from 6 studies. American Journal of Clinical Nutrition. 2009;89(5):1433-40.

46. Hall A, Bobrow E, Brooker S, Jukes M, Nokes K, Lambo J, et al. Anaemia in schoolchildren in eight countries in Africa and Asia. Public Health Nutrition. 2000;4:749-56.
47. Nikoi E, Anthamattan P. Childhood anaemia in Ghana: an examination of associated socioeconomic and health factors. African Geographical Reviews 2014;33:19-35.
48. Owusu-Agyei S, Fryauff DJ, Chandramohan D, Koram KA, Binka FN, Nkrumah FK, et al. Characteristics of severe anemia and its association with malaria in young children of the Kassena-Nankana District of northern Ghana. Am J Trop Med Hyg. 2002;67(4):371-7.
49. Ronald LA, Kenny SL, Klinkenberg E, Akoto AO, Boakye I, Barnish G, et al. Malaria and anaemia among children in two communities of Kumasi, Ghana: a cross-sectional survey. Malar J. 2006;5:105.
50. Kassebaum NJ, Jasrasaria R, Naghavi M, Wulf SK, Johns N, Lozano R, et al. A systematic analysis of global anemia burden from 1990 to 2010. Blood. 2014;123(5):615-24.
51. Yang Z, Lönnerdal B, Adu-Afarwuah S, Brown KH, Chaparro CM, Cohen RJ, et al. Prevalence and predictors of iron deficiency in fully breastfed infants at 6 mo of age: comparison of data from 6 studies. Am J Clin Nutr. 2009;89(5):1433-40.
52. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet. 2008;371(9608):243-60.
53. Ruel MT, Alderman H, Group MaCNS. Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? Lancet. 2013;382(9891):536-51.
54. Kandala NB, Madungu TP, Emina JB, Nzita KP, Cappuccio FP. Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? BMC Public Health. 2011;11:261.
55. Nesbitt RC, Gabrysch S, Laub A, Soremekun S, Manu A, Kirkwood BR, et al. Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. Int J Health Geogr. 2014;13:25.

Table 4.1: Baseline and Endline characteristics of the Ghana trial participants

	Baseline	Endline
Trial participants with a geo-coded residence (n)	1943	1781
Males (%)	992 (51.1)	900 (50.5)
Enrolment age (months), mean (SD)	19.2 (8.50)	19.3 (8.50)
Serum ferritin (µg/L), geometric mean (SD)	35.1 (3.65)	73.9 (3.58)
CRP (mg/L), mean (SD)	3.34 (4.96)	3.86 (5.10)
Maternal Education, n (%)[a,b]		
None	586 (33.5)	543 (33.0)
Any	1166 (66.5)	1105 (67.0)
Household Asset Score, n (%)[a]		
Low	866 (47.5)	817 (47.7)
High	957 (52.5)	897 (52.3)
Age of introduction of complementary foods, n (%)[a]		
≤6 months	1573 (87.3)	1475 (87.0)
>6 months	229 (12.7)	220 (13.0)
Consumption of iron fortified foods[a]		
Consumed at least 1 iron fortified product in previous 7 days prior to enrolment	4 (0.22)	4 (0.23)

[a] Measured at baseline only; reduced sample size at baseline (approximately 1825) due to incomplete surveys and "unknown" responses

[b] Total n=1752 at baseline (74 respondents were not mothers)

Table 4.2: Results from spatial Models 1-5 for log-transformed ferritin concentration[1] at baseline among 1943 young Ghanaian children (Brong-Ahafo Region, March-April 2010).

Covariates	Estimate (95% CrI) on the exponential scale[2]		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI) on the log scale	
				Spatial	Compound
1) **Child-level Model**					
Intercept	40.74	(32.41, 52.08)	3.614	0.380	0.008
Age per month			(1.836, 6.793)	(0.271, 0.549)	(0.004, 0.029)
6-23 months	1.031	(1.020, 1.042)			
24-35 months	1.043	(1.022, 1.064)			
Sex (male reference)	1.253	(1.123, 1.398)			
2) Household-level Model					
Intercept	35.91	(21.93, 57.17)	4.356	0.359	0.007
Asset score	1.082	(1.013, 1.155)	(2.087, 8.162)	(0.236, 0.559)	(0.004, 0.028)
Maternal education	0.856	(0.751, 0.976)			
Distance to health facility (km)	1.051	(0.987, 1.114)			
Distance to district centre (km)	0.999	(0.974, 1.028)			
3) Spatial Model					

Intercept	33.36	(22.93, 47.61)	4.755	0.407	0.007
Elevation (m)	1.000	(0.997, 1.004)	(2.157, 9.383)	(0.273, 0.620)	(0.004, 0.028)
Urban/built-up land (LC13)	0.832	(0.598, 1.147)			
Woody savannahs (LC8)	1.010	(0.516, 1.967)			
NDVI	1.073	(0.430, 2.593)			
NDVI*LC8	0.793	(0.429, 1.479)			
NDVI*LC13	1.448	(0.630, 3.418)			
4) Final Model					
Intercept	34.38	(21.75, 53.32)	3.786	0.347	0.007
Age per month			(1.793, 7.493)	(0.234, 0.541)	(0.004, 0.028)
6-23 months	1.033	(1.022, 1.045)			
24-35 months	1.036	(1.015, 1.058)			
Sex (male reference)	1.253	(1.119, 1.402)			
Asset score	1.067	(1.003, 1.135)			
Distance to health facility (km)	1.054	(0.993, 1.115)			
Distance to district centre (km)	0.999	(0.976, 1.026)			
Urban/built-up land (LC13)	0.814	(0.605, 1.101)			
Woody savannahs (LC8)	0.822	(0.423, 1.585)			
5) Maximal Model					
Intercept	13.64	(13.64, 47.53)	4.120	0.332	0.007
Age per month			(1.751, 8.650)	(0.219, 0.518)	(0.004, 0.028)

6-23 months	1.032	(1.021, 1.044)
24-35 months	1.040	(1.018, 1.063)
Sex (male reference)	1.238	(1.103, 1.388)
Assets	1.060	(0.994, 1.129)
Maternal education	0.898	(0.790, 1.022)
Distance to health facility (km)	1.071	(1.002, 1.140)
Distance to district centre (km)	1.000	(0.976, 1.027)
Elevation (m)	1.002	(0.998, 1.005)
Urban/built-up land (LC13)	0.911	(0.651, 1.272)
Woody savannahs (LC8)	0.852	(0.438, 1.644)
NDVI	1.153	(0.453, 2.866)
NDVI*LC8	0.831	(0.456, 1.519)
NDVI*LC13	1.396	(0.575, 3.462)

[1]Baseline serum ferritin concentration (µg/dL) corrected for baseline CRP concentration using the regression method (Namaste et al.). Corrected ferritin values were log transformed.

[2]Exponentials of posterior medians (relative risk) and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases iron levels by 5%.

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

NDVI = normalized difference vegetation index, averaged over 2010 yearly values, centred by dividing by 1000 and subtracting 4

NDVI*LC8 = interaction term between NDVI and LC=8

NDVI*LC13 = interaction term between NDVI and LC=13

Table 4.3: Results from spatial Models L4 and L5 for log-transformed ferritin concentration[1] at endline among 1781 young Ghanaian children (Brong-Ahafo Region, May-November 2010)

Covariates	**Estimate (95% CrI) on the exponential scale[2]**		**Range parameter in km (95% CrI)**	**Standard deviations of random effects (95% CrI) on the log scale**	
				Spatial	**Compound**
L4) Final Model					
Intercept	45.98	(27.25, 89.62)	4.081	0.393	0.007
Age per month			(1.555, 9.627)	(0.263, 0.593)	(0.004, 0.028)
6-23 months	1.039	(1.027, 1.050)			
24-35 months	1.005	(0.984, 1.026)			
Sex (male reference)	1.123	(1.003, 1.258)			
Asset score	0.974	(0.915, 1.037)			
Distance to health facility (km)	1.029	(0.963, 1.098)			
Distance to district centre (km)	0.999	(0.965, 1.025)			
Urban/built-up land (LC13)	0.764	(0.531, 1.059)			
Woody savannahs (LC8)	1.098	(0.556, 2.188)			
Baseline iron status	1.153	(1.081, 1.230)			
Group	1.197	(0.866, 1.652)			

Group*Baseline iron status	1.013	(0.930, 1.103)			
L5) Final Model					
Intercept	46.65	(22.00, 111.6)	4.365	0.433	0.007
Age per month			(1.565, 10.77)	(0.298, 0.633)	(0.004, 0.028)
6-23 months	1.034	(1.023, 1.046)			
24-35 months	1.008	(0.986, 1.031)			
Sex (male reference)	1.114	(0.992, 1.252)			
Asset score	0.962	(0.902, 1.027)			
Maternal education	0.887	(0.778, 1.010)			
Distance to health facility (km)	1.038	(0.958, 1.125)			
Distance to district centre (km)	1.001	(0.964, 1.031)			
Elevation (m)	1.002	(0.997, 1.006)			
Urban/built-up land (LC13)	0.711	(0.483, 1.013)			
Woody savannahs (LC8)	1.183	(0.584, 2.424)			
NDVI	0.875	(0.334, 2.316)			
NDVI*LC8	0.892	(0.461, 1.717)			
NDVI*LC13	1.237	(0.489, 3.087)			
Baseline iron status	1.149	(1.075, 1.227)			
Group	1.144	(0.819, 1.597)			
Group*Baseline iron status	1.023	(0.937, 1.118)			

[1]Endline serum ferritin concentration (μg/dL) corrected for endline CRP concentration using the regression method (Namaste et al.). Corrected ferritin values were log transformed.

[2]Exponentials of posterior medians (relative risk) and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases iron levels by 5%.

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

NDVI = normalized difference vegetation index, averaged over2010 yearly values, centred by dividing by 1000 and subtracting 4

NDVI*LC8 = interaction term between NDVI and LC=8

NDVI*LC13 = interaction term between NDVI and LC=13

Baseline iron status = serum ferritin concentration at baseline (µg/dL) corrected for baseline CRP concentration using the regression method

Group = intervention arm allocation (1=iron group, 0=no-iron group)

Group*Baseline iron status = interaction between intervention arm (1=iron group, 0=no-iron group) and baseline serum ferritin concentration corrected for baseline CRP concentration using the regression method

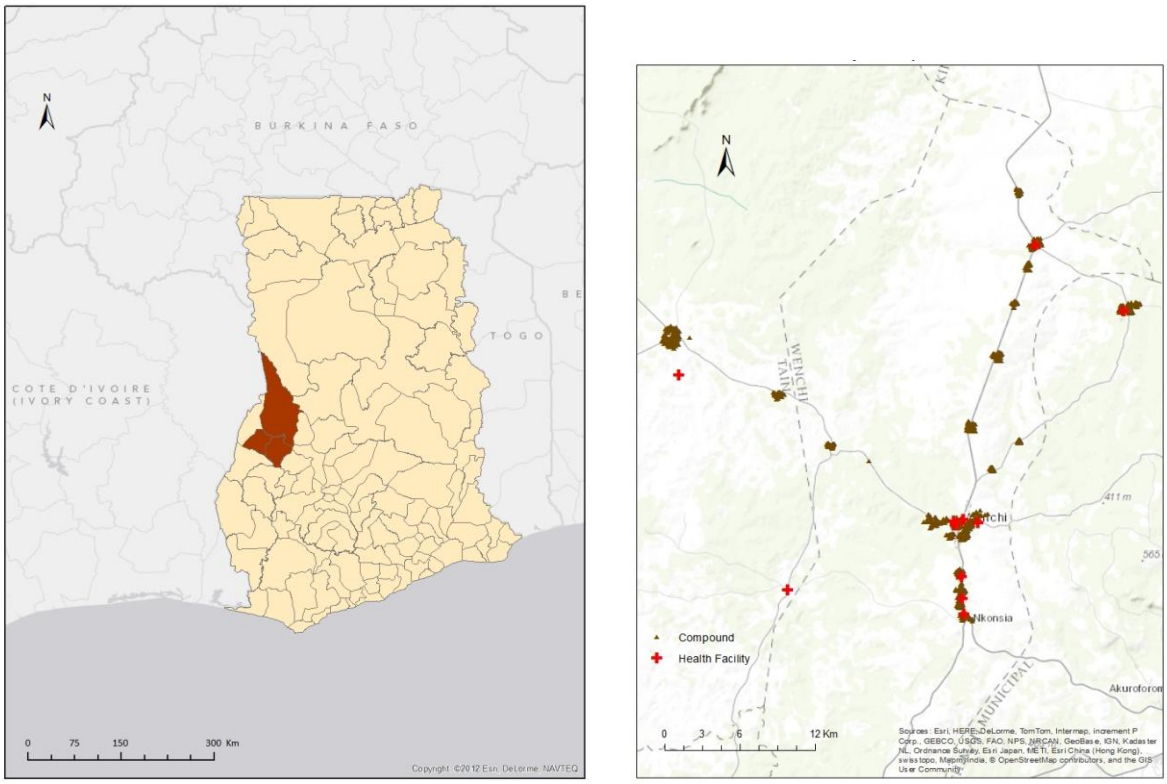
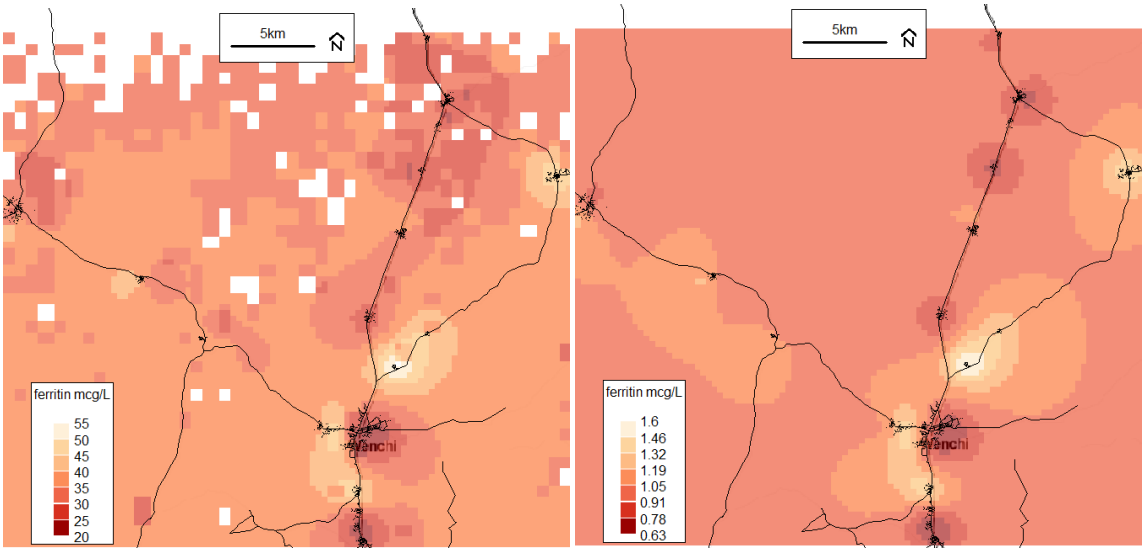
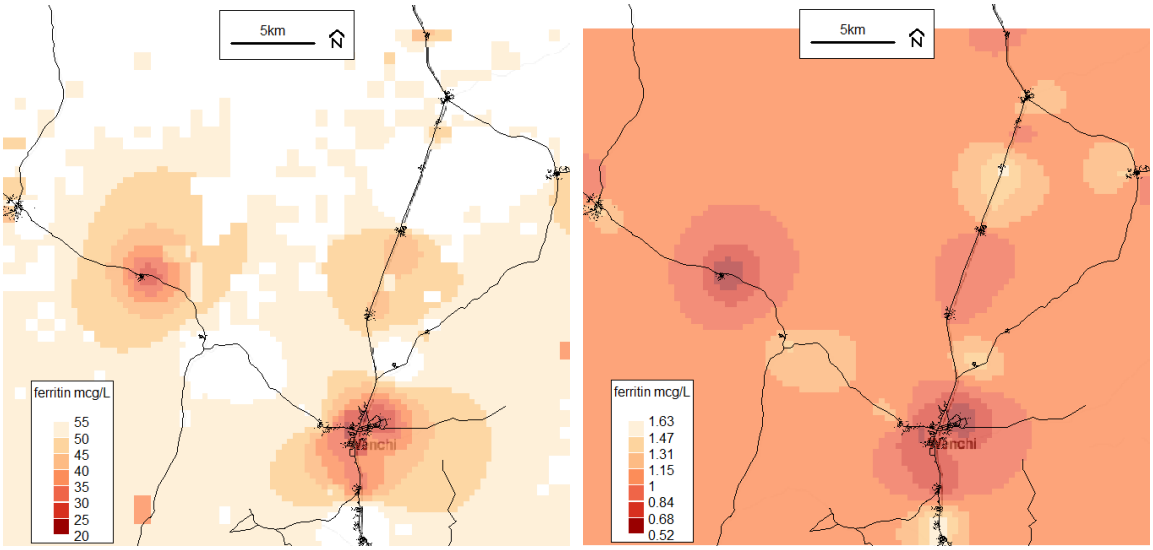


Figure 4.1: Left panel: Wenchi and Tain Districts (red) in the Brong-Ahafo Region of Ghana; Right panel: Location of Ghana study compounds (brown triangles) and health facilities (red crosses).



Figures 4.2a and 4.2b: Predicted mean raster plot from Model 4 (left), and residual spatial variation plot from Model 4 (right). Darker colour indicates lower ferritin concentration at baseline and thus higher risk of iron deficiency. Background © Stamen Design.



Figures 4.3a and 4.3b: Predicted mean raster plot from Model L4 (left), and residual spatial variation plot from Model L4 (right). Darker colour indicates lower ferritin concentration at endline and thus higher risk of iron deficiency. Background © Stamen Design.

Chapter 5
Aim 2 Manuscript

[To be formatted for submission to Malaria Journal:
http://www.malariajournal.com/authors/instructions]

5 Geo-spatial factors associated with infection risk among young children in rural Ghana

Ashley M. Aimone Phillips, Patrick E. Brown, Stanley H. Zlotkin, Donald C. Cole, Seth Owusu-Agyei

Abstract

Background: Determining the spatial patterns of infection among young children living in a malaria endemic area may provide a means of locating high risk populations who could benefit from additional resources for treatment and improved access to health care. The objective of this secondary analysis of baseline, cross-sectional data from a cluster-randomized trial among 1943 young Ghanaian children (6-35 months of age) was to determine the geo-spatial factors associated with malaria and non-malaria infection status.

Methods: Spatial analyses were conducted using a generalized linear geostatistical model with a Matern spatial correlation function and four definitions of malaria and non-malaria infection status: 1) inflammation (C-reactive protein, CRP >5 mg/L) and/or malaria parasitaemia; 2) inflammation without parasitaemia; 3) parasitaemia with fever (clinical malaria); and 4) any parasitaemia (with/without fever or inflammation). Potentially informative variables were included in the final model through five modeling steps: individual-level variables (Model 1); household-level variables (Model 2); satellite-derived spatial variables (Model 3); and a

minimal and maximal set of covariates from Models 1 to 3 (Models 4 and 5, respectively). Where applicable, iron status was defined using plasma ferritin concentration corrected for inflammation using a regression-based method.

Results: The final models indicated that children with inflammation (CRP >5 mg/L) and/or any evidence of malaria parasitaemia at baseline were more likely to be under 2 years of age, stunted, wasted, live farther from a health facility, live at a lower elevation, have less educated mothers, and have higher (corrected) ferritin concentration compared to children without inflammation or parasitaemia. Similar results were found when infection was defined as clinical malaria or parasitaemia with/without fever (definitions 3 and 4). Conversely, when infection was defined using CRP only, all covariates were non-significant with the exception of baseline iron status. In Model 5, all infection definitions that included parasitaemia demonstrated a significant interaction between normalized difference vegetation index (NDVI) and land cover type. Maps of the predicted mean and spatial random effect showed defined high- and low-risk areas that tended to coincide with elevation and cluster around villages.

Conclusions: The risk of infection among young children in a malaria endemic area may have a predictable spatial pattern that is associated with geographical characteristics such as elevation and distance to a health facility.

Keywords

Spatial, infection, malaria, children, geostatistical modeling, Bayesian inference

Background

According to the World Health Organization (WHO), the leading causes of death in children less than 5 years of age are infection-related – primarily pneumonia, diarrhea, and malaria – and approximately 45% of all deaths are associated with malnutrition [1]. Child mortality rates are highest in low- and middle-income countries (LMICs), particularly sub-Saharan Africa where the risk of death is 15 times greater than in high income regions [1]. Malnourished children are more vulnerable to infections, primarily due to compromised immune function and epithelial integrity, and inflammation [2]. For example, Muller et al. (2003) reported a positive association between malaria morbidity and the degree of protein-energy malnutrition among children in West Africa [3]. Micronutrient deficiencies also have a compromising effect on immune function, which can usually be improved through diet changes, food fortification, or supplementation [4]. However, for iron nutrition, the relationship between iron deficiency and infection risk is less clear. Evidence suggests that providing iron as supplements or through fortification to children with high infection exposure may or may not increase the risk of infection-related morbidity and mortality [5-8]. Conversely, inflammation due to infection can affect iron homeostasis [9] and the risk of iron deficiency [10], particularly in cases of prolonged or chronic infection. Therefore, assessing the risk of infection is an important first step in developing safe and effective means of administering iron to children in LMICs where iron deficiency and anaemia are prevalent.

Infection status can be assessed using biomarkers such as C-reactive protein (CRP), an acute phase protein that becomes elevated in response to the early phase of the inflammatory response (approximately 24-48 hours) [11]. The feasibility of measuring such indicators among children in a low-resource context is limited, especially at a population level, as they may require relatively large blood samples and sophisticated analytical methods with laboratory equipment. As such, there are clear advantages in identifying indicators or risk factors associated with infection in LMICs that are not invasive or costly to measure, and thus provide a more feasible means of identifying high-risk populations. This need could be addressed with geographical factors (or “geo-indicators”), as the environmental or spatial characteristics of a village or region

could provide insight into the dynamics and distribution of infection risk among children. Collecting geo-spatial data is non-invasive and less costly compared to biological measures; and they are often publicly available, which improves the access to and comparability of population-level statistics across regional and national borders.

There is mounting evidence to support the use of geographical information systems (GIS) and spatial analysis methods for conducting disease surveillance and risk analysis, assessing health system access, and informing health system planning [12-15]. In terms of infectious disease research, there are a limited number of examples where geostatistical methods have been used to investigate the spatial patterns and associated risk factors of malaria or other infections among children in LMICs [16-18]. Even fewer studies have used spatial analysis to link the geographical variation of infection with iron deficiency risk among children in low-resource settings [19]. Soares Magalhaes and colleagues. used survey data to build Bayesian geostatistical models to determine the relative contribution of parasitic infections (malaria and helminth) to the spatial variation of anaemia risk among children (≤15 years of age) in northern Angola [19]. The authors found that anaemia, *P. falciparum* and *S. haematobium* tended to cluster around inland bodies of water, and estimated that approximately 15.6% and 9.7% of the spatial variation of anaemia risk was attributable to malaria and schistosomiasis, respectively [19]. While Soares Magalhaes and colleagues provided a good starting point for the integration of infection control programs with iron supplementation, a drawback of their analyses was the use of anaemia as an indicator of iron status rather than a more specific biomarker such as ferritin concentration. There are many causes of anaemia in addition to iron deficiency [20].

Considering the bi-directional relationship between infection and iron homeostasis, the ability to describe the spatial variation of infection risk while accounting for iron status may allow us to more confidently identify areas where integrated infection and iron deficiency control programs are most needed. The objective of the current analysis was to determine the geo-spatial factors associated with malaria and non-malaria infection risk among children with varying levels of

iron sufficiency in rural Ghana. The sections that follow include a summary and interpretation of the results of this analysis, as well as a discussion of their contribution to, and implications for, global health research.

Methods

Study population

The data used in these analyses were generated from a community-based cluster randomized trial conducted in 2010 in the Wenchi and Tain Districts of the Brong-Ahafo region, a substantially rural area of Ghana [6]. At the time, there were an estimated 7.2 million cases of malaria per year in Ghana, and the prevalence of anaemia among preschool aged children was 76.1% (95% CI 73.9-78.2%) [21,22]. Briefly, the aim of the randomized trial was to determine the effect of providing iron with other micronutrients in powder form for 5 months during the rainy season (March to November) on the incidence of malaria among 1958 children aged 6-35 months (representing 1552 clusters and 22 villages) (**Figure 1**) [6]. The geographical layout of the trial area, including study compounds, health facilities, and road networks, is depicted in **Figure 2**.

Secondary measures from trial data

At baseline, biological samples were collected and analyzed for biochemical indicators of iron and infection status, including serum ferritin, C-reactive protein (CRP), and malaria parasite density. Plasma CRP was measured using an immunoturbidimetric method (QuickRead CRP, Orion Diagnostica, Espoo, Finland), and serum ferritin an enzyme immunoassay (Spectro Ferritin S-22, Ramco Laboratories Inc., Stafford, USA). Thin and thick smears were prepared for malaria parasite speciation and count via microscopy [see Zlotkin et al. (2013) for a complete description of biochemical and infection measures]. Demographic and nutrition-related information was collected at the household and individual levels, including household

assets, maternal education, feeding practices, and child body weight and length. Z-scores for weight-for-length, length-for-age were calculated using the WHO Child Growth Standards [23].

Geographical coordinates

Handheld global positioning system (GPS) units were used to collect geographical coordinates for over 95% of the study compounds (representing 1943 trial participants), as well as 22 study villages, and surrounding health facilities and road networks . The GPS coordinates were measured using the WGS 1984 coordinate system and transformed using a universal transverse Mercator (UTM) Zone 30N projection (EPSG code: 32630).

Satellite-derived data

Elevation data were downloaded from the U.S. Geological Survey (USGS) [24], with a spatial resolution of 3 arc-seconds (approximately 90 meters). Normalized Difference Vegetation Index (NDVI) data, obtained from the Land Processes Distributed Active Archive Center (LPDAAC) [25], were produced by a spectroradiometer that uses blue, red, and near-infrared reflectance to determine vegetation indices for 16-day intervals with a 250-meter spatial resolution. Land cover (LC) data (downloaded from worldgrids.org) had a spatial resolution of 500 meters, and consisted of 17 land cover classes sub-grouped into three categories: natural vegetation (11 classes), developed and mosaic land (3 classes), and non-vegetation (3 classes) [26].

Spatial modeling

The data were analyzed using generalized linear geostatistical models (GLGM) [27, 28]. Four definitions of baseline infection status served as the dependent variables: 1) inflammation (CRP >5 mg/L) and/or malaria parasitaemia; 2) inflammation (CRP >5 mg/L) without parasitaemia; 3) parasitaemia with measured concurrent fever (axillary temperature $>37.5^{0}$C) or reported history of fever within 48 hours (i.e. clinical malaria)); and 4) parasitaemia with or without concurrent

fever or history of fever. All dependent variables were binary-valued (coded as '1' for positive infection status), and analyzed using a logistic model. Geo-spatial and non-spatial variables were chosen for inclusion in the final models based on expert opinion and a review of the literature pertaining to spatial risk factors of malaria and anaemia among young children in low- and middle-income countries [18]. Variables were eligible for inclusion if they were considered to be direct or indirect antecedent factors associated with infection (e.g. elevation), and excluded if they were potential outcomes of infection (e.g. anaemia).

The models were fit using Bayesian inference via an Integrated Nested Laplace Approximation (INLA) algorithm [29]. Given the exploratory nature of the analyses, weak or uninformative priors were used for all model parameters with the exception of the Matern shape parameter which was fixed at 2. Spatial predictions were made on a 100-cell grid covering the study area. The Matern correlation, approximated by a Markov random field [30], extended an additional 3000 m in each direction. Infection probabilities, after transformation with a logit link function, were modeled as the sum of the contributions of the explanatory variables, as well as spatially correlated and compound-level random effect terms. The posterior medians of the odds of infection were computed, assuming baseline values for individual-level covariates and location-specific values for the spatial covariates. A spatially continuous (or geostatistical) model was used for the spatial random effect term, where the correlation between the log-odds of infection of two individuals was given by a Matern spatial correlation function and applied to the distance separating their respective compounds. All spatial modeling was conducted using the *glgm* function from the "geostatsp" package in R [31, 32].

For each outcome, selected variables were included in the spatial analyses in a series of modeling steps. Model 1 included individual-level variables only: baseline child age, sex, weight-for-length z-score and length-for-age z-score, and baseline iron status (ferritin concentration corrected for CRP using a regression-based method [Namaste et al., unpublished[i]]). Age in months was calculated using the reported date of birth and trial enrolment

date. The age variable was included in all models with a change point at 24 months, as this was the closest half-year to the mean age of those children who were no longer receiving breast milk (mean=26.8 months ± 5.8, n=746). Similar age variable definitions have been used in other studies of iron deficiency and anaemia in children [33, 34].

Model 2 included only household-level variables: asset score, maternal education, and distance from each compound to the nearest health facility. Household asset score was generated using a principal component analysis of 6 economic indicators (farm ownership, size and type of crops grown, type of toilet facility, and house ownership). For descriptive purposes, asset score was dichotomized at the median; however, it was modeled continuous variable. Maternal education was included as a binary variable, representing “none” (0) versus “any” (1) level of education (e.g. primary, middle, secondary or higher). Distance to the nearest health facility (an indicator of access to the health care system) was measured ‘as the bird flies' (straight-line or Euclidean distance) using the Near Table tool in ArcMap (ArcGIS 10.2, Environmental Systems Resource Institute, Redlands, California).

Five satellite-derived variables were included in Model 3: elevation, land cover type (LC), normalized difference vegetation index (NDVI), and two NDVI-LC interaction terms. Elevation across the trial area ranged from 116 to 530 meters, and values were centered by subtracting 250 meters before including the final elevation raster in the analyses. Land cover type was a discrete categorical variable consisting of 3 values: woody savannah (LC=8, n=21/1943 observations), urban and built up land (LC=13, n=243/1943 observations), and cropland/natural vegetation mosaic (LC=14, n=1679/1943 observations). In all analyses, the largest category (cropland/natural vegetation mosaic) was used as the reference. NDVI was included as an indicator of ‘greenness’, as well as a proxy for soil moisture [16], and was averaged over the year that the study was conducted (2010) in a single raster file. Averaged NDVI values had a range of 0.22-0.62. An interaction term for NDVI and LC was created by, first, using the NDVI raster to mask the LC raster except in areas where LC had a cell value of 8 (woody savannah).

The unmasked cells were then given a value of zero. The same method was also used to create the NDVI-LC interaction term for LC values of 13 (urban and built-up land). The new rasters for the interaction terms were then included in the analyses to investigate whether the association between the dependent variable (infection status) and vegetation (or soil moisture) varied across areas with or without a woody savannah or urban/built-up land cover type.

The final model (Model 4) combined selected variables from Models 1-3, including age, sex, weight-for-length z-score, length-for-age z-score, baseline iron status (serum ferritin corrected for CRP using the regression method and re-scaled by multiplying each corrected value by the inverse of the inter-quartile range), asset score, distance to the nearest health facility, and elevation. As a confirmatory modeling step, a 'maximal' model (Model 5) was also developed and included the same variables as the final model with the addition of maternal education, NDVI, LC, and the two NDVI-LC interaction terms.

Maps of predicted infection probabilities and residual spatial variation from the final model (Model 4) were plotted and overlaid with a base map of the trial area. The residual spatial variation plot represented the posterior mean of the spatial random effect, corresponding to the difference between the predicted and expected odds of infection at each location (given the spatial covariate at each location). The plots were visually compared to each other and to relevant satellite-derived maps (e.g. elevation) in order to generate potential explanations for the spatial patterns observed.

Ethics

Ethics approval for the original clinical trial was obtained from the Kintampo Health Research Centre (KHRC) Institutional Ethics Committee, the Ghana Health Service (GHS) Ethical Review Committee, the Hospital for Sick Children Research Ethics Board, and the Food and

Drugs Authority of Ghana. The secondary analysis of trial data, as well as the collection and primary analysis of geographical data, was approved by the Hospital for Sick Children and University of Toronto Health Sciences Research Ethics Boards. Informed consent was obtained from each participant's primary caregiver before screening and enrolment in the trial.

Results

Table 1 shows the baseline characteristics (biochemical measures, anthropometrics, and demographics) of 1943 children with geocoded compounds who were included in this secondary analysis (Figure 1). The mean age at enrolment was 19.2 months, with 69% (1348/1943) of participants aged below 24 months. After correcting ferritin concentration for inflammation (CRP) using the regression method, the prevalence of iron deficiency (ferritin <12 µg/L) was 21.4% (415/1943) at baseline. According to CRP measures and parasite counts, approximately one third of all children had an infection at baseline (719/1943, 37.0%). The prevalence of wasting (<-2SD for weight-for-length z-score), and stunting (<-2SD for length-for-age z-score) was 8.1% (158/1942) and 13.8% (267/1934), respectively.

The results from modeling steps 1-3 have been included in an additional file [see Additional file 1**]**. Briefly, the definition of infection that included both CRP and parasitaemia seemed to be the most sensitive to covariate associations. In Model 1, age (6-23 months), and baseline iron status were positively associated with infection (CRP>5 mg/L and/or parasitaemia); while age (24-36 months), length-for-age z-score, and weight-for-length z-score were negatively associated with infection status (Table S1). In Model 2, lower maternal education, and greater distance to the nearest health facility were associated with positive infection status (Table S2). The only satellite-derived spatial variable associated with infection in Model 3 was elevation, indicating that lower elevation corresponded with higher infection risk at baseline (Table S3).

Results from the final models (Model 4) indicated that children with inflammation (CRP >5 mg/L) and/or any evidence of malaria parasitaemia at baseline were more likely to be between 6 and 23 months of age (OR 1.03, 95% credible interval (CrI) 1.01, 1.05), approximately 10% more likely to be stunted or wasted (OR 0.91 for length-for-age z-score and 0.89 for weight-for-length z-score), live farther from a health facility (10% increased odds of infection for each km) and at a lower elevation (7% increased odd of infection for every 10 meters), and/or have higher ferritin concentration (OR 1.45, 95% CrI 1.32, 1.59) compared to children without inflammation or parasitaemia (**Table 2**). Similar results were found when infection was defined as clinical malaria or parasitaemia with/without fever (definitions 3 and 4); however, the magnitude of the association with distance to a health facility increased up to a 20% greater likelihood of infection with each km of separation. Conversely, when infection was defined using CRP only (without parasitaemia), all covariates were non-significant with the exception of baseline iron status.

The predicted infection probabilities (odds) and residual spatial variation from all final models are illustrated in **Figures 2-6**. For infection status defined using parasitaemia (**Figures 3a, 5a, 6a**), the relationship between infection risk and elevation was apparent particularly when compared to an elevation map of the study area (**Figure 7**). For infection defined using CRP only, there appeared to be well-defined high- and low-risk areas that tended to cluster around villages (**Figure 4a**). Unlike the infection definitions using parasitaemia, however, the plot of the spatial random effect for CRP only (**Figure 4b**) was similar to that of the predicted odds (**Figure 4a**), further supporting the observation that the covariates included in this final model did not explain a large amount of spatial variation in non-malaria infection.

Overall, the results from Model 5 confirmed those of Model 4, with the addition of maternal education being negatively associated with parasitaemia with or without fever (OR for “any” education 0.67, 95% CrI 0.52, 0.87) (**Table 3**). All maximal models that included parasitaemia

as part of the outcome definition also had a significant NDVI-LC interaction term, suggesting that outside of urbanized and built-up areas, each 0.1 “unit” increase in greenness (with units ranging between 0.22 and 0.62) was associated with increases in the odds of infection of greater than 40%.

Most models demonstrated significant spatial random effects, indicating that there was residual variation in the odds of baseline infection across the study area, particularly when infection was defined as inflammation without parasitaemia. Comparatively, the compound random effects tended to be small with narrow 95% credible intervals (CrI), indicating relatively low variability in infection risk between compounds. The range parameter from each model indicated that the distance at which the inter-variable relationships started to decay (decreased covariance) ranged from 3.45 (95% CrI 1.56, 6.97) to 7.25 (95% CrI 2.82, 14.4) kilometers in the final models (Model 4 set), and 4.84 (95% CrI 1.98, 9.83) to 6.75 (95% CrI 2.15, 15.9) kilometers in the maximal models (Model 5 set).

Discussion

To the authors’ knowledge, the geostatistical analyses presented herein are the first to demonstrate spatial relationships for the risk of malaria and non-malaria infection, using standard and novel definitions, among children living in a malaria endemic area with varying levels of iron status. In particular, elevation and distance to the nearest health facility were consistently associated with infection when it was defined using parasitaemia, either alone or in combination with CRP or fever. For example, in a final model, children with inflammation (CRP >5 mg/L) and/or malaria parasitaemia at baseline were more likely to live farther from a health facility and at a lower elevation. Access to the health care system is generally considered

to be a positive predictor of health, and this relationship is supported by other studies in malaria endemic areas [16, 35, 36]. Assuming that malaria was the largest contributor to the prevalence of infection in this study's population of Ghanaian children, the inverse relationship observed between elevation and infection status is consistent with other studies conducted in malaria endemic areas showing a lower prevalence of malaria among populations living at higher elevations [37-40]. The prediction raster plots for the final models (Model 4 set) especially illustrate this relationship when compared to an elevation map of the study area (**Figure 7**). The association between malaria and elevation is related to temperature, as the early stages of parasite development are sensitive to temperature and will be delayed or inhibited in colder environments, which are found at higher altitudes [41].

In terms of individual-level risk factors, children with parasitaemia (with or without high CRP) were more likely to be under 2 years of age, be stunted or wasted, and/or have higher ferritin concentration at baseline compared to children without infection. Infants tend to be at higher risk of infection due to immature immune systems [42]; particularly at later stages of infancy when they begin to explore their environment, which increases the risk of exposure to pathogens. Both stunting (low length for age) and wasting (low weight for length) have also been associated with a higher risk of infection-related morbidity and mortality among children less than 5 years of age in low- and middle-income countries [43, 44]. The only individual-level variable that remained significant across all definitions of infection in both Model sets 4 and 5 was baseline iron status. The relationship in all models was also positive, indicating that those with higher serum ferritin concentrations were more likely to have high CRP and/or parasitaemia. This is not surprising considering the well-known up-regulating effect of infection or inflammation on acute phase proteins like ferritin. Although baseline ferritin values were corrected for the effect of inflammation (using the regression method), the only biomarker available for this was CRP. During the early phase of the inflammatory response, CRP reaches its peak concentration within 24-48 hours [11]. When the concentration of CRP declines, ferritin tends to remain elevated. Therefore, for more complete ferritin correction, an additional acute phase protein that corresponds to the late phase of the inflammatory response, such as alpha-1-

acid glycoprotein (AGP), is needed [11]. In this case, it is possible that the prevalence of inflammation was underestimated, resulting in incomplete correction of ferritin and residual confounding.

Similar to Model 4, among the Model 5 set, infection defined using parasitaemia was the most informative in terms of identifying environmental and non-environmental relationships. Maternal education was negatively associated with infection status, which agrees with the generally reported finding that parental schooling has a positive influence on child health and nutrition status [45]. The significant NDVI-LC interaction term may be explained by the effect of land cover and land use on the survival and breeding behaviors of mosquitoes. Relationships between infection risk and vegetation type have been reported by others investigating the spatial risk factors of malaria in Ghana [46] and Indonesia [47]. For example, Krefis and colleagues (2011) found a lower incidence of malaria among children living in forested areas of rural Ghana (RR = 0.53), while those living in close proximity to cultivation had a higher risk of malaria [46]. Forest-type vegetation may be less likely to collect water where mosquitoes could breed and more easily infect those who live nearby [48]. On the other hand, areas that have been cleared or cultivated may be more likely to have standing water due to irrigation, certain topographical characteristics (e.g. slope), or poor drainage [37, 39, 49]. Appawu and colleagues (2004) demonstrated higher malaria transmission rates in irrigated communities compared to non-irrigated areas of the Kassena Nankana District of northern Ghana, where the land is primarily used for subsistence farming [49]. In the western highlands of Kenya, Cohen and colleagues found that households with confirmed malaria cases tended to be closer to areas with high wetness indices (predicted water accumulation), which were generated using hydrologic modeling of land surface water flow [37]. In the present analyses, the interaction between vegetation or greenness (NDVI) and “urbanized or built-up” land cover type may reflect the differentiation between forested and cultivated land, and thus corresponding propensities for water to accumulate and create breeding grounds for Anopheles mosquitoes.

The significant spatial random effects observed in all models suggest that geographical distribution may be important to consider when assessing infection risk in a population. This was especially apparent when the final model outputs were plotted and compared to an elevation map, further demonstrating the utility of GIS and spatial analysis in exploring and communicating population health risks and characteristics. In some cases, particularly for infection defined using CRP only, mapping the spatial random effect from the final model suggested that the factors included in the analysis may not have fully explained the variation observed. While a larger sample size and/or geographical coverage may have allowed us to gain additional insight into potential sources of spatial variation, this may also increase the risk of a Type I error. An additional limitation pertains to CRP and its inability to capture the late phase of the inflammatory response (i.e. after 48 hours), as this may have lead to underestimation of non-malaria infection prevalence and thus an incomplete picture of spatial variability.

Similar to the spatial random effects, the model range remained relatively constant across models though was quite a bit smaller compared to other spatial analyses of malaria prevalence in sub-Saharan Africa. Ashton and colleagues used spatial modeling with a Bayesian framework to assess the spatial variation of malaria (*Plasmodium falciparum and vivax*) among 5,914 school children in Oromia Regional State, Ethiopia [50]. They described range as the distance at which similarities in climatic factors and ecology would be expected, and found that it was approximately 45 kilometers in the *P. falciparum* model [50]. Although the outcome assessed by Ashton and colleagues was similar to that of the Ghana trial (*P. falciparum* parasitaemia), the range from each study may have been less comparable due to differences in key study characteristics. These included the measurement methods used (e.g. enzyme-linked immunosorbent assay versus microscopy in the Ghana trial for assessing malaria seroprevalence), the covariates included in the spatial models (e.g. environmental factors only versus a combination of environmental, individual- and household-level factors in the Ghana trial), and the size of the study area (284,500 km^2 in Ethiopia versus 3,200 km^2 in Ghana).

Unlike the spatial random effects, the compound random effect observed in all models tended to be relatively small, suggesting that there was low additional variability in the outcome (infection status) across compounds. Since a compound may have consisted of more than one household, we expected to find some spatial clustering of the outcome at the compound level. Potential explanations for why this was not observed include: 1) the compounds within a village were in close proximity to each other, resulting in clustering at the village level rather than between compounds; and 2) the small number of observations per compound reduced the opportunity for the outcome to cluster within compounds. Considering that the average cluster size was 1.3, we suggest that the latter explanation was the more likely scenario.

A potential limitation of the present analyses was the use of straight-line (Euclidean) distance to estimate proximity to a health facility rather than an indicator of access by road such as network distance. While, network distance may have more appropriately accounted for travel distance by vehicle or bicycle, it was not possible to calculate due to incomplete or missing vector information (e.g. miss-aligned junctions, missing or disconnected road segments). Another study conducted in the Brong-Ahafo region of Ghana, by Nesbitt and colleagues (2014), compared different measures of travel impedance to estimate access to delivery care [51]. The authors encountered similar challenges with calculating network distance, and found that it was as informative as straight-line distance for determining geographical access in this area of rural Ghana [49]. Considering these findings, we suggest that the use of Euclidean distance in the present analyses could be justified. An additional limitation of the analyses presented here is the cross-sectional nature of the data, as it does not allow us to infer causality or eliminate the risk of reverse causality.

Conclusions

Determining the spatial dynamics of infection among children in a malaria endemic area, without the use of invasive and costly measurement methods, may provide a means of locating high risk populations and identifying geographical areas where treatment and prevention strategies should be focused. Furthermore, considering the relationship between inflammation and iron homeostasis, our maps of infection risk could also inform the geographical distribution of iron deficiency risk, or at least help to identify areas where extra caution should be used when providing iron interventions to infants and young children. Future research should include longitudinal analyses to examine the co-variation in geo-spatial factors associated with infection status over time, and to further explore the potential importance of baseline effects.

List of abbreviations used

AGP	Alpha-1-acid glycoprotein
CBC	Complete blood count
CRP	C-reactive protein
GHS	Ghana Health Service
GLGM	Generalized linear geostatistical Model
GPS	Global positioning system
IGBP	International Geosphere Biosphere Programme
INLA	Integrated Nested Laplace Approximation
KHRC	Kintampo Health Research Centre
LC	Land cover type
LMIC	Low- and Middle-Income Countries
LPDAAC	Land Processes Distributed Active Archive Center
NDVI	Naturalized Difference Vegetation Index
UTM	Universal Transverse Mercator

USGS United States Geological Survey

WHO World Health Organization

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AAP, SHZ and SOA conducted the original trial in Ghana. AAP and SOA coordinated the acquisition of geographical data. AAP conceived and conducted the secondary analysis with substantial contribution from PEB. DCC and SHZ were also involved in the conception and design of the secondary analysis, and interpretation of data. AAP drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge Seeba Amenga-Etego and the members of the GIS team at the Kintampo Health Research Centre in (Kintampo, Ghana) who collected GPS data that were used in the secondary analysis. Funding for this manuscript was provided by AAP's Doctoral Research Award from the Canadian Institutes of Health Research (CIHR).

Endnotes

[i] Namaste S, Rohner R, Suchdev P, Kupka R, Mei Z, Bhushan N, et al. Approaches to address the association between inflammation and ferritin in children and women. Unpublished.

References

1. WHO: **Children: reducing mortality.** In *Fact sheet No 178*. World Health Organization2014.
2. Katona P, Katona-Apte J: **The interaction between nutrition and infection.** *Clin Infect Dis* 2008, **46:**1582-1588.
3. Müller O, Garenne M, Kouyaté B, Becher H: **The association between protein-energy malnutrition, malaria morbidity and all-cause mortality in West African children.** *Trop Med Int Health* 2003, **8:**507-511.
4. Raiten DJ, Sakr Ashour FA, Ross AC, Meydani SN, Dawson HD, Stephensen CB, Brabin BJ, Suchdev PS, van Ommen B, Group IC: **Inflammation and Nutritional Science for Programs/Policies and Interpretation of Research Evidence (INSPIRE).** *J Nutr* 2015, **145:**1039S-1108S.
5. Sazawal Z, Black, RE., Ramsan, M., Chwaya, HM., Stoltzfus, RJ., Dutta, A ., et al.: **Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: community-based, randomised, placebo-controlled trial.** *Lancet* 2006, **367:**133-143.
6. Zlotkin S, Newton S, Aimone AM, Azindow I, Amenga-Etego S, Tchum K, Mahama E, Thorpe KE, Owusu-Agyei S: **Effect of iron fortification on malaria incidence in infants and young children in Ghana: a randomized trial.** *JAMA* 2013, **310:**938-947.
7. Veenemans J, Milligan P, Prentice AM, Schouten LR, Inja N, van der Heijden AC, de Boer LC, Jansen EJ, Koopmans AE, Enthoven WT, et al: **Effect of supplementation with zinc and other micronutrients on malaria in Tanzanian children: a randomised trial.** *PLoS Med* 2011, **8:**e1001125.
8. Soofi S, Cousens S, Iqbal SP, Akhund T, Khan J, Ahmed I, Zaidi AK, Bhutta ZA: **Effect of provision of daily zinc and iron with several micronutrients on growth and morbidity among young children in Pakistan: a cluster-randomised trial.** *Lancet* 2013, **382:**29-40.
9. Spottiswoode N, Duffy PE, Drakesmith H: **Iron, anemia and hepcidin in malaria.** *Front Pharmacol* 2014, **5:**125.

10. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, Ganz T, Kaplan J: **Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization.** *Science* 2004, **306:**2090-2093.
11. Thurnham DI, McCabe GP: **Influence of infection and inflammation on biomarkers of nutritional status with an emphasis on vitamin A and iron.** In *Report: Priorities in the assessment of vitamin A and iron status in populations, Panama City, Panama, 15-17 September 2010.* Edited by Organization WH. Geneva: World Health Organization; 2012: 63-80
12. Root ED, Lucero M, Nohynek H, Anthamatten P, Thomas DS, Tallo V, Tanskanen A, Quiambao BP, Puumalainen T, Lupisan SP, et al: **Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children.** *Proc Natl Acad Sci U S A* 2014, **111:**3520-3525.
13. Nykiforuk C, Flaman L: **Exploring the utilization of geographic information systems in health promotion and public health.** In *Technical report #08-001*. Edmonton, AB, Canada: School of Public Health, University of Alberta; 2008.
14. Song P, Zhu Y, Mao X, Li Q, An L: **Assessing spatial accessibility to maternity units in Shenzhen, China.** *PLoS One* 2013, **8:**e70227.
15. Masters SH, Burstein R, Amofah G, Abaogye P, Kumar S, Hanlon M: **Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis.** *Soc Sci Med* 2013, **93:**147-154.
16. Samadoulougou S, Maheu-Giroux M, Kirakoya-Samadoulougou F, De Keukeleire M, Castro MC, Robert A: **Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso.** *Parasit Vectors* 2014, **7:**350.
17. Musenge E, Vounatsou P, Kahn K: **Space-time confounding adjusted determinants of child HIV/TB mortality for large zero-inflated data in rural South Africa.** *Spat Spatiotemporal Epidemiol* 2011, **2:**205-217.
18. Aimone AM, Perumal N, Cole DC: **A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: the case of anaemia and malaria.** *International Journal of Health Geographics* 2013, **12:**1-13.

19. Soares Magalhães RJ, Langa A, Pedro JM, Sousa-Figueiredo JC, Clements AC, Vaz Nery S: **Role of malnutrition and parasite infections in the spatial variation in children's anaemia risk in northern Angola.** *Geospat Health* 2013, **7:**341-354.
20. Balarajan Y, Ramakrishnan U, Ozaltin E, Shankar AH, Subramanian SV: **Anaemia in low-income and middle-income countries.** *Lancet* 2011, **378:**2123-2135.
21. WHO: **World Malaria Report 2009.** Geneva, Switzerland: World Health Organization; 2009.
22. WHO: **Worldwide prevalence of anaemia 1993-2005.** In *WHO Global Database on Anaemia*. Geneva, Switzerland: World Health Organization; 2008.
23. WHO/UNICEF: **WHO child growth standards and the identification of severe acute malnutrition in infants and children: A Joint Statement by the World Health Organization and the United Nations Children's Fund.** Geneva, Switzerland: World Health Organization; 2009.
24. **United States Geological Survey** [https://lta.cr.usgs.gov]
25. **Land Processes Distributed Active Archive Center** [https://lpdaac.usgs.gov]
26. **Worldgrids — a public repository and a WPS for global environmental layers** [http://worldgrids.org]
27. Diggle PJ, Moyeed RA, Tawn JA: **Model-based geostatistics.** *Applied Statistics* 1998, **47:**299-350.
28. Diggle PJ, Ribeiro PJ: *Model-based geostatistics.* New York: Springer-Verlag; 2006.
29. Rue H, Martino S: **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society: Series B* 2009, **71:**319-392.
30. Lindgren F, Rue H, Lindstrom J: **An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.** *Journal of the Royal Statistical Society (Series B, Statistical Methodology)* 2011, **73, Part 4:**423-498.
31. Brown PE: **Model-based geostatistics the easy way.** *Journal of Statistical Software* 2015, **63:**1-24.
32. R: **R: A Language and Environment for Statistical Computing.** 3.2.2 edition. Vienna, Austria: R Foundation for Statistical Computing; 2015.

33. Adish AA, Esrey SA, Gyorkos TW, Johns T: **Risk factors for iron deficiency anaemia in preschool children in northern Ethiopia.** *Public Health Nutrition* 1999, **2:**243-252.

34. Jonker FA, Calis JC, van Hensbroek MB, Phiri K, Geskus RB, Brabin BJ, Leenstra T: **Iron status predicts malaria risk in Malawian preschool children.** *PLoS One* 2012, **7:**e42670.

35. Magalhães RJ, Langa A, Sousa-Figueiredo JC, Clements AC, Nery SV: **Finding malaria hot-spots in northern Angola: the role of individual, household and environmental factors within a meso-endemic area.** *Malar J* 2012, **11:**385.

36. Kadobera D, Sartorius B, Masanja H, Mathew A, Waiswa P: **The effect of distance to formal health facility on childhood mortality in rural Tanzania, 2005-2007.** *Glob Health Action* 2012, **5:**1-9.

37. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, Wilson ML: **Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands.** *Malar J* 2008, **7:**40.

38. Drakeley CJ, Carneiro I, Reyburn H, Malima R, Lusingu JP, Cox J, Theander TG, Nkya WM, Lemnge MM, Riley EM: **Altitude-dependent and -independent variations in Plasmodium falciparum prevalence in northeastern Tanzania.** *J Infect Dis* 2005, **191:**1589-1598.

39. Brooker S, Clarke S, Njagi JK, Polack S, Mugo B, Estambale B, Muchiri E, Magnussen P, Cox J: **Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya.** *Trop Med Int Health* 2004, **9:**757-766.

40. Reid H, Vallely A, Taleo G, Tatem AJ, Kelly G, Riley I, Harris I, Henri I, Iamaher S, Clements AC: **Baseline spatial distribution of malaria prior to an elimination programme in Vanuatu.** *Malar J* 2010, **9:**150.

41. Hay SI, Omumbo JA, Craig MH, Snow RW: **Earth observation, geographic information systems and Plasmodium falciparum malaria in sub-Saharan Africa.** *Adv Parasitol* 2000, **47:**173-215.

42. M'Rabet L, Vos AP, Boehm G, Garssen J: **Breast-feeding and its role in early development of the immune system in infants: consequences for health later in life.** *J Nutr* 2008, **138:**1782S-1790S.

43. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, Mathers C, Rivera J: **Maternal and child undernutrition: global and regional exposures and health consequences.** *Lancet* 2008, **371:**243-260.

44. Jones KD, Berkley JA: **Severe acute malnutrition and infection.** *Paediatr Int Child Health* 2014, **34 Suppl 1:**S1-S29.

45. Ruel MT, Alderman H, Group MaCNS: **Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition?** *Lancet* 2013, **382:**536-551.

46. Krefis AC, Schwarz NG, Nkrumah B, Acquah S, Loag W, Oldeland J, Sarpong N, Adu-Sarkodie Y, Ranft U, May J: **Spatial analysis of land cover determinants of malaria incidence in the Ashanti Region, Ghana.** *PLoS One* 2011, **6:**e17905.

47. Anthony RL, Bangs MJ, Hamzah N, Basri H, Purnomo, Subianto B: **Heightened transmission of stable malaria in an isolated population in the highlands of Irian Jaya, Indonesia.** *Am J Trop Med Hyg* 1992, **47:**346-356.

48. Hightower AW, Ombok M, Otieno R, Odhiambo R, Oloo AJ, Lal AA, Nahlen BL, Hawley WA: **A geographic information system applied to a malaria field study in western Kenya.** *Am J Trop Med Hyg* 1998, **58:**266-272.

49. Appawu M, Owusu-Agyei S, Dadzie S, Asoala V, Anto F, Koram K, Rogers W, Nkrumah F, Hoffman SL, Fryauff DJ: **Malaria transmission dynamics at a site in northern Ghana proposed for testing malaria vaccines.** *Trop Med Int Health* 2004, **9:**164-170.

50. Ashton RA, Kefyalew T, Rand A, Sime H, Assefa A, Mekasha A, Edosa W, Tesfaye G, Cano J, Teka H, et al: **Geostatistical modeling of malaria endemicity using serological indicators of exposure collected through school surveys.** *Am J Trop Med Hyg* 2015, **93:**168-177.

51. Nesbitt RC, Gabrysch S, Laub A, Soremekun S, Manu A, Kirkwood BR, Amenga-Etego S, Wiru K, Höfle B, Grundy C: **Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana.** *Int J Health Geogr* 2014, **13:**25.

Table 5.1: Baseline characteristics of the Ghana trial participants

Trial participants with a geo-coded residence (n)	1943
Males (%)	992 (51.1)
Age at enrollment (months), mean (SD)	19.2 (8.5)
Serum ferritin (µg/L), geometric mean (SD)	35.1 (3.65)
Infection status	
C-reactive protein (mg/L), mean (SD)	3.34 (4.96)
Parasite density (count/µL), geometric mean (SD)	3003.0 (5.35)
Inflammation and/or parasitaemia[a] , n (%)	719 (37.0)
Inflammation without parasitaemia[b] , n (%)	272 (14.0)
Parasitaemia with fever[c] , n (%)	150 (7.72)
Parasitaemia[d], n (%)	447 (23.0)
Anthropometric status[e]	
Weight-for-length z-score, mean (SD)	-0.63 (0.97)
Length-for-age z-score, mean (SD)	-0.81 (1.21)
Maternal Education, n (%)[f]	
None	586 (33.5)
Any	1166 (66.5)
Household Asset Score, n (%)[g]	
Low	866 (47.5)
High	957 (52.5)
Distance to the nearest health facility (kilometers), mean (SD)	2.57 (2.99)

[a]Inflammation and/or parasitaemia = CRP > 5 mg/L and/or any malaria parasitaemia;

[b]Inflammation without parasitaemia = CRP > 5 mg/L without malaria parasitaemia;

[c]Parasitaemia with fever = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

[d]Parasitaemia = any malaria parasitaemia (with/without fever)

[e]Measured at baseline; z-scores estimated using the WHO Child Growth Standards [23]

[f]Measured at baseline only; total n=1752 (74 respondents were not mothers, 117 missing due to incomplete surveys)

[g] Measured at baseline only; reduced sample size (approximately 1825) due to incomplete surveys and "unknown" responses

Table 5.2: Results from the final model of geospatial and non-spatial risk factors for baseline infection status[1] among 1943 Ghanaian children (Brong-Ahafo Region, March-April 2010)

Covariates	Odds ratios (95% CrI)		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia					
Intercept	0.417	(0.248, 0.715)	4.647	0.489	0.007
Age per month			(1.926, 9.273)	(0.294, 0.830)	(0.004, 0.028)
6-23 months	1.029	(1.008, 1.051)			
24-35 months	0.954	(0.919, 0.991)			
Sex (male reference)	1.084	(0.882, 1.331)			
Length-for-age z-score	0.912	(0.832, 0.999)			
Weight-for-length z-score	0.899	(0.805, 1.003)			
Asset score	1.025	(0.915, 1.149)			
Distance to health facility (km)	1.098	(1.003, 1.207)			
Elevation (m)	0.993	(0.987, 0.998)			
Baseline iron status	1.447	(1.324, 1.589)			
2) Inflammation without parasitaemia					

Intercept	0.145	(0.082, 0.261)	7.011	0.453	0.007
Age per month			(2.609, 15.39)	(0.233, 0.885)	(0.004, 0.028)
6-23 months	0.991	(0.965, 1.019)			
24-35 months	0.963	(0.912, 1.015)			
Sex (male reference)	1.149	(0.877, 1.506)			
Length-for-age z-score	1.008	(0.894, 1.136)			
Weight-for-length z-score	0.918	(0.796, 1.0575)			
Asset score	0.974	(0.845, 1.123)			
Distance to health facility (km)	0.909	(0.818, 0.999)			
Elevation (m)	0.999	(0.993, 1.004)			
Baseline iron status	1.206	(1.109, 1.312)			
3) Parasitaemia with fever					
Intercept	0.032	(0.012, 0.078)	7.249	0.836	0.007
Age per month			(2.823, 14.42)	(0.461, 1.536)	(0.028, 0.004)
6-23 months	1.036	(0.997, 1.077)			
24-35 months	0.925	(0.856, 0.994)			
Sex (male reference)	1.173	(0.804, 1.712)			
Length-for-age z-score	0.942	(0.793, 1.115)			
Weight-for-length z-score	0.815	(0.660, 1.003)			
Asset score	1.034	(0.828, 1.291)			
Distance to health facility (km)	1.163	(0.994, 1.371)			

Elevation (m)	0.991	(0.982, 0.999)			
Baseline iron status	1.317	(1.197, 1.452)			
4) Parasitaemia					
Intercept	0.197	(0.100, 0.386)	3.459	0.669	0.007
Age per month			(1.558, 6.966)	(0.420, 1.060)	(0.004, 0.028)
6-23 months	1.052	(1.025, 1.079)			
24-35 months	0.971	(0.930, 1.014)			
Sex (male reference)	1.007	(0.789, 1.285)			
Length-for-age z-score	0.875	(0.783, 0.976)			
Weight-for-length z-score	0.905	(0.793, 1.033)			
Asset score	1.058	(0.917, 1.220)			
Distance to health facility (km)	1.195	(1.064, 1.354)			
Elevation (m)	0.992	(0.985, 0.999)			
Baseline iron status	1.259	(1.160, 1.371)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (binary): 1 = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours), 0 = any malaria parasitaemia without concurrent fever or history of reported fever;

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range

Table 5.3: Results from the maximal model of geospatial and non-geospatial risk factors for baseline infection status[1] among 1943 Ghanaian children (Brong-Ahafo Region, March-April 2010)

Covariates	Odds ratios (95% CrI)		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia					
Intercept	0.218	(0.101, 0.470)	4.841	0.416	0.007
Age per month			(1.977, 9.825)	(0.243, 0.731)	(0.004, 0.028)
6-23 months	1.029	(1.007, 1.051)			
24-35 months	0.966	(0.929, 1.005)			
Sex (male reference)	1.096	(0.889, 1.352)			
Length-for-age z-score	0.911	(0.830, 1.000)			
Weight-for-length z-score	0.889	(0.795, 0.994)			
Asset score	1.016	(0.905, 1.141)			
Maternal education	0.796	(0.634, 1.001)			
Distance to health facility (km)	1.140	(1.042, 1.250)			
Elevation (m)	0.995	(0.989, 1.000)			
Urban/built up land (LC13)	0.958	(0.531, 1.711)			

Woody savannahs (LC8)	0.819	(0.272, 2.384)			
NDVI	0.604	(0.122, 2.891)			
NDVI*LC8	0.622	(0.246, 1.559)			
NDVI*LC13	5.281	(1.109, 26.64)			
Baseline iron status	1.418	(1.296, 1.560)			
2) Inflammation without parasitaemia					
Intercept	0.122	(0.046, 0.310)	6.582	0.498	0.008
Age per month			(2.495, 14.20)	(0.252, 0.983)	(0.004, 0.029)
6-23 months	0.991	(0.964, 1.019)			
24-35 months	0.969	(0.916, 1.022)			
Sex (male reference)	1.152	(0.876, 1.517)			
Length-for-age z-score	1.008	(0.892, 1.138)			
Weight-for-length z-score	0.914	(0.791, 1.055)			
Asset score	0.979	(0.845, 1.134)			
Maternal education	1.150	(0.846, 1.575)			
Distance to health facility (km)	0.906	(0.802, 1.014)			
Elevation (m)	0.999	(0.993, 1.005)			
Urban/built up land (LC13)	1.040	(0.533, 2.006)			
Woody savannahs (LC8)	1.024	(0.168, 4.187)			
NDVI	1.162	(0.167, 7.147)			
NDVI*LC8	0.820	(0.260, 3.045)			

NDVI*LC13	1.052	(0.191, 6.647)			
Baseline iron status	1.207	(1.107, 1.316)			
3) Parasitaemia with fever					
Intercept	0.004	(0.001, 0.019)	6.751	0.697	0.008
Age per month			(2.150, 15.89)	(0.328, 1.429)	(0.004, 0.030)
6-23 months	1.043	(1.002, 1.086)			
24-35 months	0.929	(0.858, 0.999)			
Sex (male reference)	1.274	(0.864, 1.882)			
Length-for-age z-score	0.907	(0.758, 1.081)			
Weight-for-length z-score	0.795	(0.639, 0.985)			
Asset score	1.015	(0.803, 1.284)			
Maternal education	0.942	(0.623, 1.438)			
Distance to health facility (km)	1.316	(1.118, 1.575)			
Elevation (m)	0.996	(0.987, 1.005)			
Urban/built up land (LC13)	1.247	(0.272, 5.094)			
Woody savannahs (LC8)	2.215	(0.323, 11.04)			
NDVI	0.114	(0.003, 3.022)			
NDVI*LC8	0.956	(0.213, 4.496)			
NDVI*LC13	40.77	(1.244, 2107)			
Baseline iron status	1.336	(1.212, 1.477)			
4) Parasitaemia					

Intercept	0.084	(0.032, 0.220)	4.841	0.514	0.008
Age per month			(1.581, 11.21)	(0.274, 0.953)	(0.004, 0.030)
6-23 months	1.053	(1.026, 1.081)			
24-35 months	0.981	(0.938, 1.025)			
Sex (male reference)	1.018	(0.794, 1.306)			
Length-for-age z-score	0.873	(0.779, 0.976)			
Weight-for-length z-score	0.897	(0.784, 1.025)			
Asset score	1.052	(0.910, 1.217)			
Maternal education	0.669	(0.515, 0.872)			
Distance to health facility (km)	1.255	(1.128, 1.414)			
Elevation (m)	0.995	(0.988, 1.001)			
Urban/built up land (LC13)	0.728	(0.290, 1.731)			
Woody savannahs (LC8)	0.781	(0.226, 2.460)			
NDVI	0.364	(0.038, 3.024)			
NDVI*LC8	0.603	(0.222, 1.620)			
NDVI*LC13	11.55	(1.249, 130.5)			
Baseline iron status	1.241	(1.141, 1.353)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (binary): 1 = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours), 0 = any malaria parasitaemia without concurrent fever or history of reported fever;

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

NDVI = normalized difference vegetation index, averaged over2010 yearly values, centered by dividing by 1000 and subtracting 4

NDVI8 = interaction term between NDVI and LC=8

NDVI13 = interaction term between NDVI and LC=13

Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range

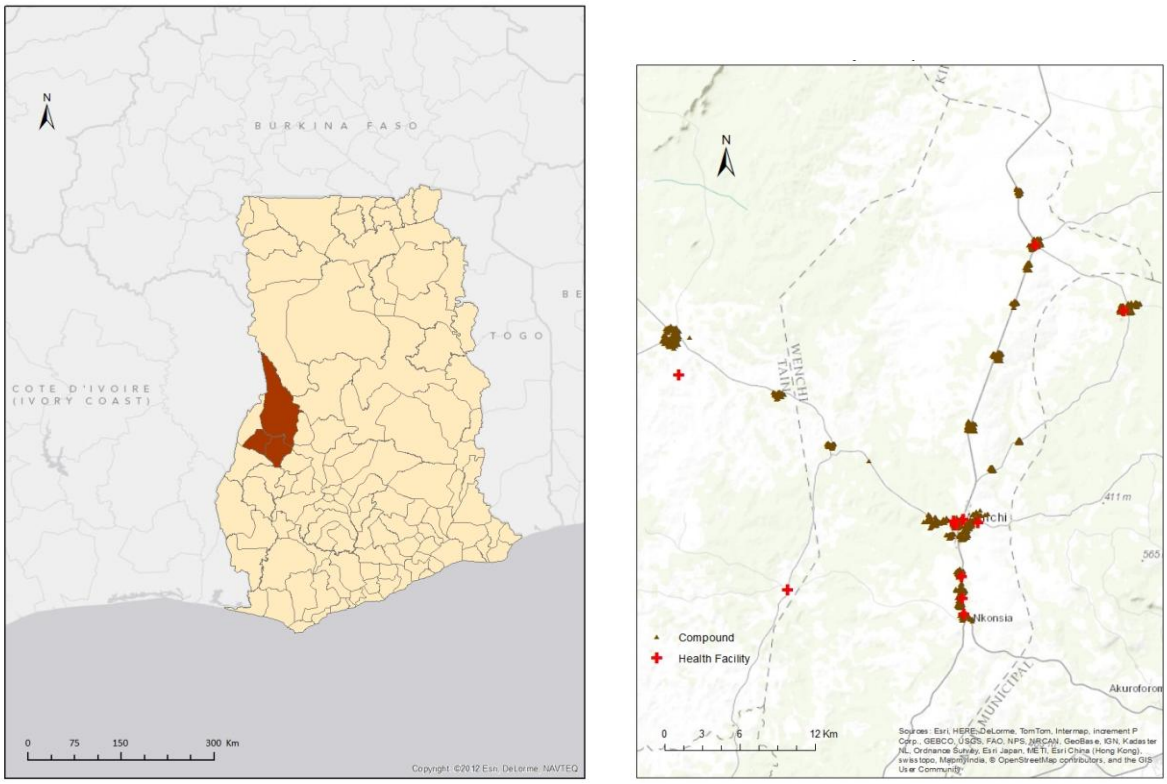
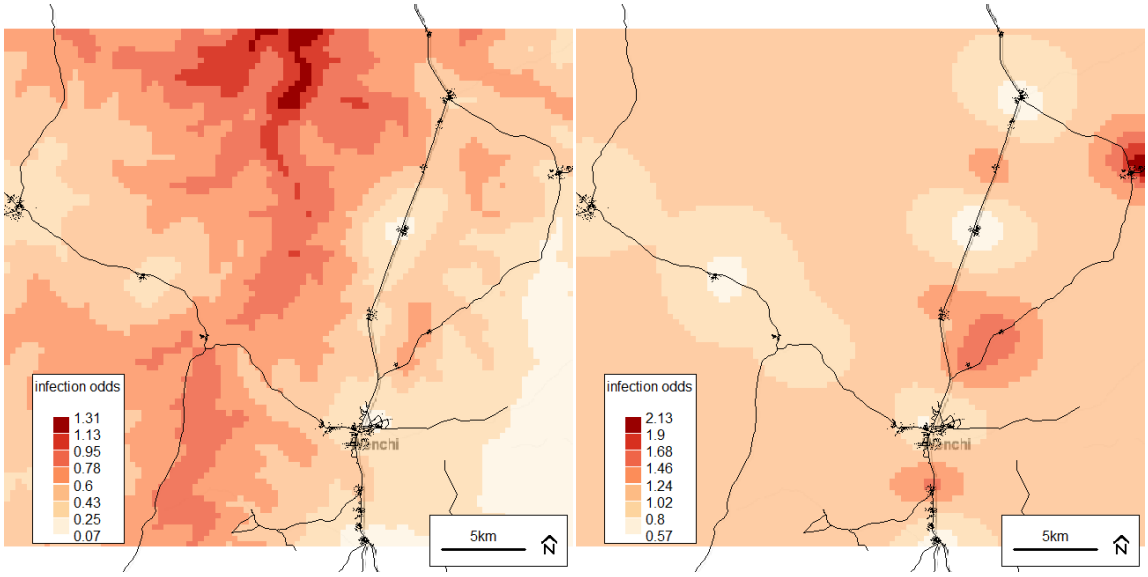
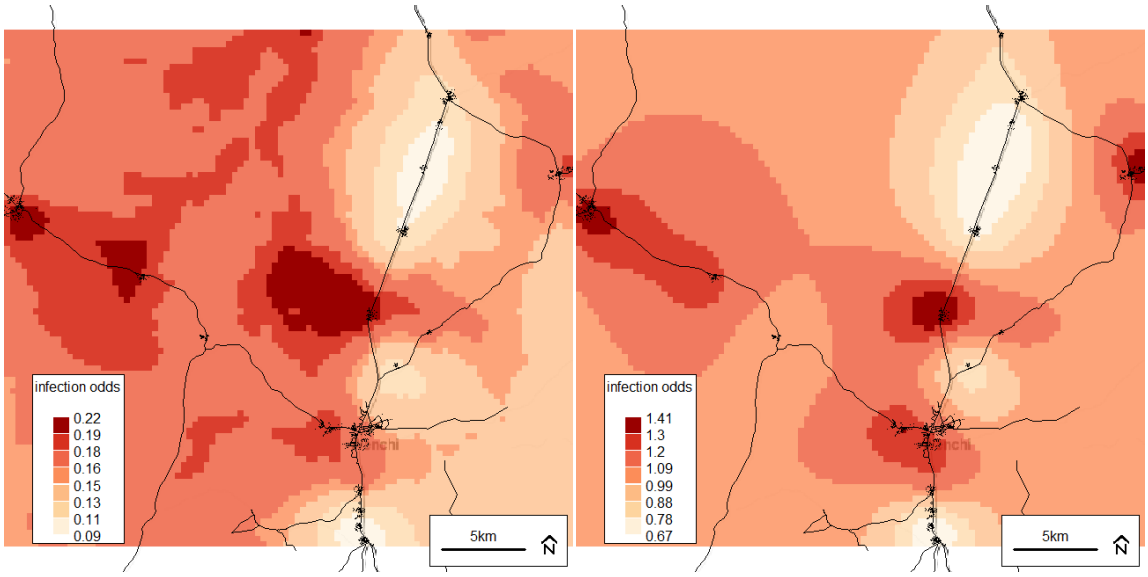


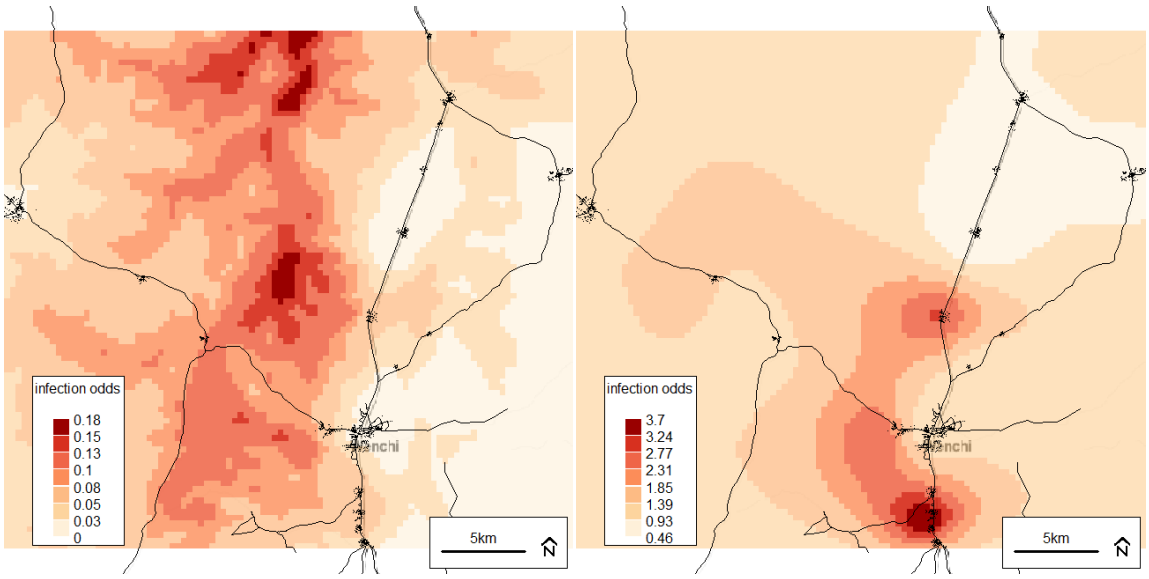
Figure 5.1: Left panel: Wenchi and Tain Districts (red) in the Brong-Ahafo Region of Ghana; Right panel: Location of Ghana study compounds (brown triangles) and health facilities (red crosses).



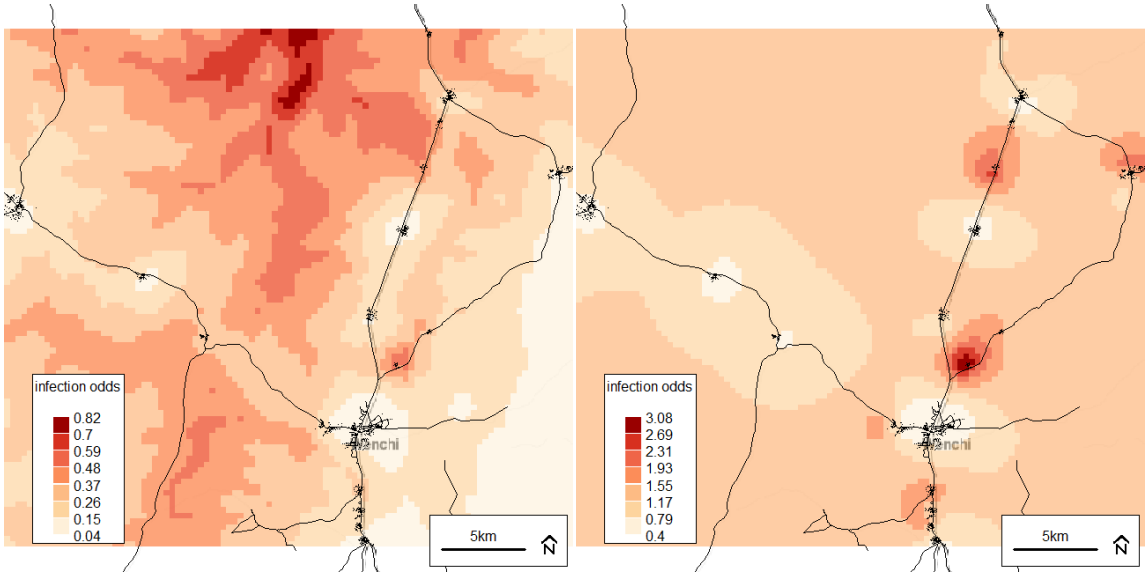
Figures 5.2a and 5.2b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final model (Model 4). Darker colour indicates higher risk (odds) of inflammation (CRP>5 mg/L) *and/or any* malaria parasitaemia at baseline. Background © Stamen Design.



Figures 5.3a and 5.3b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final model (Model 4). Darker colour indicates higher risk (odds) of inflammation (CRP > 5 mg/L) *without* malaria parasitaemia at baseline. Background © Stamen Design.



Figures 5.4a and 5.4b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final model (Model 4). Darker colour indicates higher risk (odds) of malaria parasitaemia *with* concurrent fever (axillary temperature >37.5^{0}C - or history of reported fever within 48 hours) at baseline. Background © Stamen Design.



Figures 5.5a and 5.5b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final (Model 4). Darker colour indicates higher risk (odds) of malaria parasitaemia with or without fever at baseline. Background © Stamen Design.

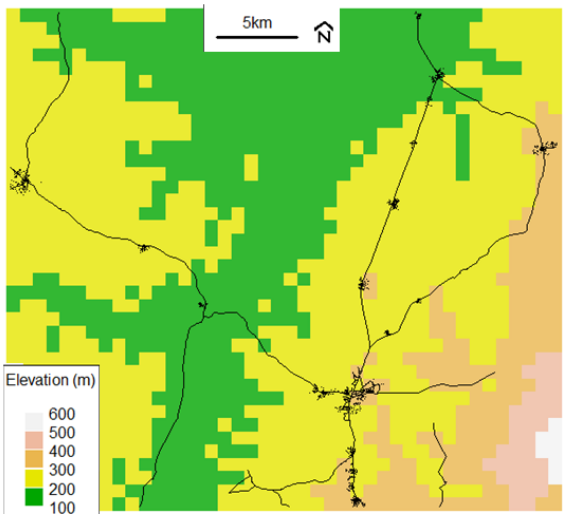


Figure 5.6: Plot of elevation changes across the study area. Green colour indicates lower elevation. Black dots represent trial compounds. Lines represent major roads.

Chapter 6
Aim 3 Manuscript

[Note: To be formatted for submission to the Journal of Epidemiology and Community Health: http://jech.bmj.com/site/about/guidelines.xhtml]

6 Impact of iron home-fortification with micronutrient powders on the geo-spatial patterns of infection risk among young children in rural Ghana

Ashley M. Aimone Phillips, Patrick E. Brown, Seth Owusu-Agyei, Stanley H. Zlotkin, Donald C. Cole

Abstract

Background: Determining the effect of iron-containing micronutrient powders on the spatial patterns of infection among children with varying levels of iron sufficiency in a malaria endemic area may provide a means of locating high risk populations where integrated infection and iron deficiency control programs may be most needed. The objective of this secondary analysis was to determine the geo-spatial factors associated with infection status among Ghanaian children at the end of a 5-month iron intervention trial.

Methods: Data from a cluster-randomized trial were analyzed using a generalized linear geostatistical model with a Matern spatial correlation function and four definitions of infection: 1) inflammation (C-reactive protein, CRP >5 mg/L) and/or malaria parasitaemia; 2) inflammation without parasitaemia; 3) parasitaemia with fever; and 4) any parasitaemia (with/without fever or inflammation). Analyses were stratified by intervention group and

compared to a non-spatial model to assess the independent effects of the iron intervention and spatial correlation.

Results: Overall, the by-group and combined-group analyses both showed that baseline infection status may have been the most important predictor of endline infection risk, particularly when infection was defined using parasitaemia (with or without inflammation).

In the No-iron group, age above 24 months and weight-for-length z-score at baseline were associated with high CRP at endline. In the combined-group spatial analyses, the risk of elevated CRP and/or parasitaemia at endline was negatively associated with household asset score, and positively associated with baseline infection status. Maps of the predicted mean and spatial random effect showed a defined low-risk area around the District centre, regardless of how infection was defined.

Conclusions: In a clinical trial setting of iron home-fortification, where all children receive bed nets and access to malaria treatment, there may be predictable geographical variation in the risk of infection with distinct low-risk areas, particularly around municipal centres. Overall, our findings emphasize the importance of considering geographical distribution when assessing infection risk in a paediatric population.

Keywords

Spatial, infection, malaria, children, geostatistical modelling, Bayesian inference

Background

According to the World Health Organization (WHO), the leading causes of death in children less than 5 years of age are infection-related. Child mortality rates are highest in low- and middle-income countries, particularly sub-Saharan Africa where the risk of death is 15 times greater than in high income regions (1). Approximately 45% of all deaths are associated with malnutrition (1), including micronutrient deficiencies. Iron deficiency is the most common nutritional disorder worldwide (2). It is estimated to account for 20,854 deaths and 2.2 million disability-adjusted life years lost per year among children less than five years of age, most of whom live in low- and middle-income countries (LMICs) (3). Iron deficient children are more vulnerable to infections, primarily due to compromised immune function (4). Conversely, inflammation due to infection can have adverse effects on iron homeostasis (5) and increase the risk of iron deficiency (6). Iron status can usually be improved through diet changes, food fortification, or supplementation. To add to this complexity, evidence from a large randomized trial conducted in Pemba, Zanzibar in 2003 indicated that in malaria endemic areas, supplementing young children with iron may increase their risk of malaria and infection-related morbidity and mortality, particularly if they are iron replete (7). Therefore, assessing the risk of infection is an important component of developing safe and effective means of administering iron to children in low- and middle-income countries (LMICs), particularly where malaria is prevalent.

Infection status can be assessed using biomarkers such as C-reactive protein (CRP), an acute phase protein that becomes elevated in response to the early phase of the inflammatory response (approximately 24-48 hours) (8). The feasibility of measuring such indicators at the population level in a low-resource context is limited, as they require blood samples and additional analyses with specialized equipment. As such, there is a need to identify indicators or risk factors associated with infection in LMICs that are not invasive or costly to measure, and thus provide a more feasible means of identifying high-risk populations.

This need could be addressed with geographical factors (or "geo-indicators"), as the environmental or spatial characteristics of a village or region could provide additional insight into the dynamics and distribution of infection among children that informs treatment needs (9-11), and helps implementers and policy makers in LMICs determine where health services should be distributed in order to make efficient use of limited resources. Collecting geo-spatial data is non-invasive and tends to be less costly compared to biological measures; and they are often publicly available, which improves the access to and comparability of population-level statistics across regional and national borders. There is mounting evidence supporting the use of geographical information systems (GIS) and spatial analysis methods for conducting disease surveillance and risk analysis, assessing health system access, and informing health system planning (12-15). In terms of paediatric nutrition and infectious disease research, there are a limited number of examples where geostatistical methods have been used to investigate the spatial patterns and associated risk factors of iron deficiency or malaria and other infections among children in LMICs (16-18). Even fewer studies have used spatial analysis to link the geographical variation of infection with iron deficiency risk among children in low-resource settings (11). Magalhaes et al. used survey data to build Bayesian geostatistical models to determine the relative contribution of parasitic infections (malaria and helminth) to the spatial variation of anaemia risk among children (≤15 years of age) in northern Angola (11). The authors found that anaemia, *P. falciparum* and *S. haematobium* tended to cluster around inland bodies of water, and estimated that approximately 15.6% and 9.7% of the spatial variation of anaemia risk was attributable to malaria and schistosomiasis, respectively (11). While these studies provided a good starting point for informing the integration of infection control programs with micronutrient supplementation, their drawbacks include: 1) the use of Hb or anaemia as a primary outcome, which is not a specific biomarker of iron status like ferritin concentration; 2) the assessment of infection risk without accounting for iron status, which could have an impact on immune function; 3) the predominant use of cross sectional data, thus reducing the ability to make causal inferences or avoid the risk of reverse causality; and 4) the lack of data pertaining to the effect of providing iron (e.g. in the form of supplements or fortification) on the spatial distribution or geo-indicators of infection risk. Considering the bi-directional relationship between infection and iron homeostasis, the ability to describe the spatial variation of infection risk while accounting for the effect of time, iron status, and/or iron

interventions may allow us to more confidently determine how and where to implement integrated infection and iron deficiency control programs.

We previously conducted a spatial analysis to determine the geo-spatial factors associated with baseline infection status among iron deficient and sufficient children upon enrolment in a randomized iron intervention trial in rural Ghana (19). The results of these analyses suggested that the risk of infection (defined using a combination of parasitaemia and CRP concentration) may be related to elevation, and distance to the nearest health facility. In the current analyses, we have expanded the dataset to include endline data to determine whether the geo-spatial factors of infection status among Ghanaian children at the end of the randomized iron intervention trial were consistent with those observed at baseline, and whether these associations were influenced by baseline iron status or the iron intervention itself.

Methods

Study population

The data used in these analyses were generated from a community-based cluster randomized trial conducted in 2010 in a rural area of Ghana (Brong-Ahafo Region) (20). At the time, Ghana had an estimated 7.2 million cases of malaria per year, and the prevalence of anaemia among preschool aged children was 76.1% (95% CI 73.9-78.2%) (21, 22). Briefly, the aim of the clinical trial was to determine the effect of providing iron in powder form with other micronutrients (micronutrient powders) for 5 months on the incidence of malaria among 1958 children aged 6-35 months (representing 1552 clusters and 22 villages) (20). The geographical layout of the study area has been described elsewhere (19).

Measures from trial data

Biological samples collected at baseline and endline were analyzed for biochemical indicators of iron and infection status, including serum ferritin, C-reactive protein (CRP), and malaria parasite density. Plasma CRP and serum ferritin were measured using an immunoturbidimetric method (QuickRead CRP, Orion Diagnostica, Espoo, Finland) and an enzyme immunoassay (Spectro Ferritin S-22, Ramco Laboratories Inc., Stafford, USA), respectively. Thin and thick smears were prepared for malaria parasite speciation and count via microscopy. Malaria screening was also performed on a weekly basis throughout the intervention period using a rapid diagnostic antigen test (Paracheck Pf®). If a child had a fever within the past 48 hours or an axillary temperature >37.5^{0}C, a blood sample was drawn and analyzed in the field via antigen test, and in the lab using microscopy. Antigen test results were used to determine treatment needs, and parasite density was combined with fever information to calculate clinical malaria incidence (episode counts) [see Zlotkin et al. (2013) for a complete list of biochemical and infection measures]. Demographic and nutrition-related information was collected at baseline at the household and individual levels, including household assets, maternal education, feeding practices, and child body weight and length. Z-scores for weight-for-length, length-for-age were calculated using the WHO Child Growth Standards (23).

Geographical data

Approximately 24 months after completion of the clinical trial (August-November 2012), geographical coordinates for the study villages (n=22) and compounds (n=1539), health facilities (n=22), and road networks were collected using handheld global positioning system (GPS) units. The GPS coordinates were measured using the WGS 1984 coordinate system and transformed using a universal transverse Mercator (UTM) Zone 30N projection (EPSG code: 32630).

Elevation data (SRTM DEM Version 3) were downloaded from the United States Geological Survey (USGS) (24), which is a government agency for the natural sciences, drawing on data such as the National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) data sets. The data file had a resolution of 3 arc-seconds (approximately 90 meters). Normalized Difference Vegetation Index (NDVI) data were downloaded from the Land Processes Distributed Active Archive Center (LPDAAC) (25). The NDVI data was produced by the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument (product MOD13Q1), which uses blue, red, and near-infrared reflectance to determine vegetation indices for 16-day intervals with 250-meter spatial resolution. The MODIS land cover type product (MOD12Q1) was downloaded from worldgrids.org and consisted of 17 land cover classes (defined by the International Geosphere Biosphere Programme or IGBP), including natural vegetation (11 classes), developed and mosaic land (3 classes), and non-vegetation (3 classes). Land cover data were derived from yearly output from the Terra and Aqua MODIS instruments at a resolution of 500 meters (26).

Variable Preparation

Age in months was calculated using the reported date of birth and trial enrolment date. For the age variable we set a change point at 24 months, as this was the closest half-year to the mean age of those children who were no longer receiving breast milk (mean=26.8 months ± 5.8, n=746). A change point of 24 months has also been used in other studies of iron deficiency and anaemia in children (27, 28). Household asset score was generated using a principal component analysis of 6 economic indicators, including farm ownership, size and type of crops grown (e.g. cash crop for sale in the market or for household consumption), type of toilet facility, and house ownership (e.g. own vs. rent). For descriptive purposes, asset score was dichotomized at the median. Maternal education was included as a dichotomous variable representing “none” versus “any” education. Baseline iron status was defined as serum ferritin concentration corrected for inflammation (baseline CRP) using the regression method (29), and re-scaled by multiplying the corrected values by the inter-quartile range. Distance to the nearest health facility was included as an indicator of access to the health care system, and was measured ‘as the bird files' (or

Euclidean) using the Near Table tool in ArcMap (ArcGIS 10.2, Environmental Systems Resource Institute, Redlands, California).

All satellite-derived data were downloaded as global datasets and cropped according to the geographical boundaries of the trial area. Elevation had a range of 116-530 meters, and values were centered by subtracting 250 before including the raster in the analyses. Land cover type (LC) consisted of 3 values, representing woody savannahs (LC=8, n=21/1943 observations), urban and built up land (LC=13, n=243/1943 observations), and cropland/natural vegetation mosaic (LC=14, n=1679/1943 observations). NDVI was included as a proxy for soil moisture (18). Since vegetation has the propensity to change over time due to such factors as climate change, urban development, or rural expansion, all NDVI data that were available for Ghana during the year that the trial was conducted (2010) were collected and averaged in a single raster file. Averaged NDVI values had a range of 0.22-0.62. Two NDVI-LC interaction terms were created by centering the average NDVI raster values (NDVI / 1000 – 4), and adjusting the resolution of the LC raster to match that of the NDVI raster. The NDVI raster was then used to mask the LC raster except where LC=8 (these cells were given a value of zero). This created a new raster that could be used to investigate whether the association between the dependent variable and vegetation (or soil moisture) varied across areas with or without a woody savannah land cover type. The same method was applied for the NDVI-LC interaction term with LC values of 13 (urban and built-up land).

Spatial modelling

The data were analyzed using generalized linear geostatistical models (GLGM) (30, 31) with four definitions of endline infection status as the dependent variables: 1) inflammation (CRP >5 mg/L) with/without malaria parasitaemia; 2) inflammation (CRP >5 mg/L) without parasitaemia; 3) parasitaemia with measured concurrent fever (axillary temperature >37.5^{0}C) or reported history of fever within 48 hours (i.e. clinical malaria)); and 4) parasitaemia with or

without concurrent fever or history of fever. All dependent variables were binary-valued (coded as '1' for positive infection status at endline), with the exception of the third definition (parasitaemia with fever) which was a count variable representing the number of new clinical malaria episodes during the intervention period. Geo-spatial and non-spatial variables were chosen for inclusion in the final models based on expert opinion and a review of the literature pertaining to spatial risk factors of malaria and anaemia among young children in low- and middle-income countries (16). Variables were eligible for inclusion if they were considered to be direct or indirect antecedent factors associated with infection (e.g. elevation), and excluded if they were outcomes of infection (e.g. anaemia).

The spatial models were fit using Bayesian inference via an Integrated Nested Laplace Approximation (INLA) algorithm (32). Flat uninformative priors were used for all model parameters with the exception of the Matern shape parameter which was fixed at 2. Spatial predictions were made on a 100-cell grid covering the study area, and the Matern correlation approximated by a Markov random field (33) extended an additional 3 km in each direction. Infection probabilities, after transformation with a logit or log link function (for binary or count outcomes, respectively), were modelled as the sum of the contributions of the explanatory variables and a spatially correlated and compound-level random effect terms. A spatially continuous (or geostatistical) model was used for the spatial random effect term, with the correlation between the log-odds or log-risk of infection of two individuals given by a Matern spatial correlation function applied to the distance separating their respective compounds. All spatial modelling was conducted using the *glgm* function from the "geostatsp" package in R (34, 35).

The preliminary modelling steps for the baseline spatial analyses have been described elsewhere (19). Briefly, selected variables were included in the baseline spatial analyses in five modelling steps. Model 1 included individual-level variables only (e.g. age and sex), Model 2 included household-level variables (e.g. asset score and maternal education), and Model 3 included study

area-level satellite-derived variables (e.g. elevation and normalized difference vegetation index (NDVI)). The final model (Model 4) consisted of a combination of selected variables from Models 1-3. A 'maximal' model (Model 5) was also developed as a confirmatory modelling step and included the same variables as the final model with the addition of maternal education, NDVI, LC, and the NDVI-LC interaction terms.

Unlike baseline, only the final model (Model 4) was included in the present endline analyses, though with three additional sub-modelling steps: 1) No-iron group only; 2) Iron group only; 3) both groups combined. The interventions groups were analyzed separately in order to differentiate the effect of time and iron treatment on baseline associations observed previously. The variables included in the by-group models included: baseline age in months, sex, baseline z-scores for weight-for-length and length-for-age, asset score, distance from each compound to the nearest health facility, elevation, baseline iron status (baseline serum ferritin concentration corrected for baseline CRP using the regression method, and re-scaled using the inverse of the inter-quartile range), and baseline infection status (according the definition of the dependent variable). The combined-group model included the same variables as the by-group models with the addition of group allocation, and two interaction terms for group-baseline iron status and group-baseline infection status. In addition to the effect of time and iron treatment, the effect of space was also explored by re-analyzing the combined-group model using a non-spatial generalized linear mixed model with a Bayesian framework. The non-spatial model was analyzed using the *INLA* package in R (32).

Maps of predicted infection probabilities and residual spatial variation from the combined-group models were plotted and visually compared with a base map of the trial area and coordinate points for the compound locations. Predicted infection probabilities were computed as the posterior means of the odds or risk of infection, assuming baseline values for individual-level covariates and location-specific values for the spatial covariates. The residual spatial variation plot represents the posterior mean of the spatial random effect, corresponding to the difference

between the predicted and expected odds or risk of infection at each location (given the spatial covariate at each location).

Ethics

Approval for the original clinical trial was obtained from the Kintampo Health Research Centre (KHRC) Institutional Ethics Committee, the Ghana Health Service (GHS) Ethical Review Committee, the Hospital for Sick Children Research Ethics Board, and the Food and Drugs Board of Ghana. Approval for conducting the secondary analysis of trial data, as well as the collection and primary analysis of geographical data was obtained from the Hospital for Sick Children and University of Toronto Health Sciences Research Ethics Boards.

Results

Baseline and endline characteristics of the study sample are presented in **Table 6.1**, including biochemical, infection, and anthropometric measures, and demographics. A total of 1780 trial participants were included in the endline analyses, representing those with geo-coded residences who provided blood samples at endline and had both baseline and endline CRP and parasitaemia values. For the infection outcome defined as parasitaemia with fever, there were 1939 observations included in the Poisson regression analysis, which corresponded to the number of children with geo-coded compounds who had at least one recorded follow-up visit during the intervention period, and thus contributed data to the malaria count outcome. The mean age at enrolment was 19.3 months, with 69% (1230/1780) of participants aged below 24 months.

The proportion of children with evidence of infection (elevated CRP and/or parasitaemia) increased from baseline to endline (37.0% at baseline versus 41.6% at endline). This increase may have been driven primarily by a greater prevalence of parasitaemia at endline (27.1% versus 23.0% at baseline), though the difference between iron intervention groups was small

(26.3% versus 28.0% in the No-iron and Iron groups, respectively). The prevalence of non-malaria inflammation (CRP>5 mg/L without parasitaemia) remained relatively stable over the course of the trial; however, the proportion in the Iron group was slightly higher at endline (7.6% in the Iron group versus 6.6% in the No-iron group). After correcting ferritin concentration for inflammation (CRP) using the regression method (29), the prevalence of iron deficiency (ferritin <12 µg/L) was 21.4% (415/1943) at baseline and 9.21% (164/1781) at endline (70/886 in the Iron group, 94/895 in the No-iron group).

Results from the by-group spatial analyses indicated that the risk of elevated CRP (without parasitaemia) at endline among children in the No-iron group was positively associated with baseline weight-for-length z-score (OR 1.30, 95% credible interval (CrI) 1.06, 1.60), and negatively associated with age between 24 and 36 months (OR 0.92, 95% CrI 0.84, 1.006) (**Table 6.2**). This suggests that among children who received MNPs without iron for 5 months, those with a higher weight-for-length z-score (less wasted) or at the lower end of the 24-35 month age range at baseline had increased odds of non-malaria inflammation or infection at endline of 30% and 8%, respectively. When parasitaemia was added to the outcome definition, the relationship with age above (24-36 months) was reversed (OR for definition 4: 1.07, 95% CrI 1.03, 1.11) and baseline infection status became a significant factor (OR for definition 4: 2.86, 95% CrI 1.97, 4.17). The latter association indicated that children in the No-iron group with parasitaemia (with or without CRP or fever) at baseline were up to 186% more likely to have parasitaemia at endline. In the Iron group, only baseline infection status was positively associated with endline status when defined as inflammation and/or parasitaemia (OR 2.27, 95% CrI 1.66, 3.11) or parasitaemia with/without fever (OR 2.54, 95% CrI 1.75, 3.68) (**Table 6.3**).

In the combined-group spatial analyses, the risk of elevated CRP and/or parasitaemia at endline was negatively associated with household asset score, and positively associated with baseline infection status (OR for definition 1: 0.88, 95% CrI 0.78, 0.98) (**Table 6.4**). This suggests that children from wealthier households had a 12.5% decreased odds of infection at endline,

regardless of whether they received MNPs with or without iron. Similar to the Iron group analyses, baseline infection status was significantly and positively associated with the corresponding endline outcome when defined as inflammation and/or parasitaemia (OR 1.84, 95% CrI 1.36, 2.50) or parasitaemia with/without fever (2.75, 95% CrI 1.91, 3.95).

The predicted mean and residual spatial variation from all combined-group spatial models are illustrated in **Figures 6.1-6.4**. The maps show a defined low-risk area around the District centre, regardless of how infection was defined. Conversely, the location of high-risk areas seems to vary across models, though consistencies can be seen between **Figures 6.1** and **6.4** where the geographical distribution of predicted infection probabilities roughly approximated elevation. The maps depicting residual spatial variation (**Figures 6.1a, 6.2a, 6.3a, 6.4a**) also show defined high- and low-risk areas, indicating that a large amount of spatial variation was not explained by the variables included in the analyses and further supporting the large spatial random effects observed across all spatial models. Contrary to spatial variation, the compound random effects in all spatial models tended to be small with narrow 95% credible intervals (CrI), indicating low variability in infection risk between compounds. The range parameter from each model indicated that the distance at which the inter-variable relationships started to decay (decreased covariance) ranged from 6.54 (95% CrI 2.04, 15.1) to 9.46 (95% CrI 2.39, 23.6) kilometers in the No-iron group model, 7.29 (95% CrI 2.49, 17.718.0) to 12.6 (95% CrI 4.32, 27.3) kilometers in the Iron group model, and 7.34 (95% CrI 2.93, 15.8) to 10.8 (95% CrI 8.77, 25.5) kilometers in the combined-group model.

The combined-group non-spatial analyses showed similar results in terms of non-malaria inflammation or infection (no association); however additional covariate relationships were revealed for infection definitions that included parasitaemia that were not present in the spatial analyses (**Table 6.5**). For example, parasitaemia with or without elevated CRP (definition 1) was positively associated with distance to the nearest health facility (OR 1.04, 95% CrI 1.01, 1.08), as was clinical malaria (OR 1.06, 95% CrI 1.03, 1.09). Both findings indicate that children who lived farther from a health facility had higher odds of infection (4-5% for every kilometer). Asset score was positively associated with parasitaemia plus fever (definition 3: OR

1.13, 95% CrI 1.04, 1.23), suggesting that higher household wealth was a risk factor for infection. Elevation had very weak associations with parasitaemia with/without inflammation or fever (definitions 1 and 4), which may have been spurious and were difficult to interpret.

Comparison to baseline infection risk

The factors associated with endline infection risk in both the by-group and combined-group models were not fully consistent with those observed at baseline, which included elevation and distance to a health facility, as well as individual-level variables such as age, length-for-age z-score, and iron status (19). This suggests that the effect of time and/or the intervention may have influenced the dynamics of infection risk across the study population.

At endline, the most consistent variable associated with the outcome was baseline infection status, which further emphasizes the importance of assessing infection risk upon implementing iron intervention strategies. The impact of the intervention was also evident when the plots of the predicted means of infection risk were compared between baseline and endline. At endline, there were larger and more defined low-risk areas across the study site for all infection definitions, particularly around the District centre. When infection was defined using CRP only, some of the high- and low-risk areas at baseline seemed to be reversed at endline. Similar to baseline, the plots of residual spatial variation at endline were generally not a close resemblance to their corresponding predicted mean plots, with the exception of CRP, suggesting that the covariates included in the final models did not explain a large amount of spatial variation in non-malaria infection at both baseline and endline. The ranges of the combined-group models were larger and more variable than those of the final baseline model. For example, the range for the model of infection defined as parasitaemia with/without fever was 3.46 (95% CrI 1.56, 6.97) kilometers at baseline and 7.85 (95% CrI 3.38, 16.1) kilometers at endline. The largest change in range from baseline to endline was observed for infection defined using CRP only, which was 7.01 (95% CrI 2.61, 15.4) kilometers at baseline and 10.77 (95% CrI 8.77, 25.5) kilometers at

endline. Unlike range, the compound random effect remained relatively small across all infection models from baseline to endline.

Discussion

The analyses presented herein are the first to explore the spatial variation and associated risk factors of infection, defined using both inflammatory and parasitic biomarkers, among children in a malaria endemic area (rural Ghana) after a randomized iron intervention trial. Overall, both the by-group and combined-group analyses showed that baseline infection status may have been the most important predictor of endline infection risk, particularly when infection was defined using parasitaemia (with or without inflammation). What is also interesting to note is the consistency of this relationship across intervention groups, as current evidence might lead one to expect that providing micronutrient powders (MNPs) with iron to children with malaria parasitaemia would increase their risk of subsequent parasitic infections to a greater extent than those who receive MNPs without iron (7). Another factor to consider is that all children in both intervention groups were monitored and treated for identified clinical malaria episodes (parasitaemia plus fever) throughout the trial. As shown in **Table 6.1**, most parasitaemia at baseline was asymptomatic (without fever), which may have reflected chronic infection or treatment failure of a previous infection. In either case, these asymptomatic parasitic infections would not have been treated as clinical malaria episodes, and thus may have persisted throughout the intervention period of the trial, regardless of whether iron was received.

Stratifying the analyses by group provided a window into potential interactions by separating the effect of time (No-iron group) from the effect of treatment (Iron group). In the No-iron group (**Table 6.2**), children with a higher weight-for-length z-score at baseline were 31% more likely to have high CRP (>5 mg/L) at endline (OR 1.32, 95% CrI 1.06, 1.60). This positive association was not expected, given that malnourished children are often found to be at

increased risk of infection (4). In this case, the relationship appeared in both instances when infection was defined using CRP (with or without parasitaemia), which becomes elevated during the first 48 hours of the inflammatory response (8). Since the intervention period was 5 months long, it is unlikely that anthropometric measures at baseline influenced inflammatory status at endline. Rather, it is more likely that the association was influenced by another factor that was not included in the present analyses and also associated with the intervention, since weight-for-length z-score was not a significant factor in the Iron group models (**Table 6.3**). High CRP was also inversely associated with age in the No-iron group, though only for children between 24 and 36 months (OR 0.93, 95% CrI 0.84, 1.00). Assuming that elevated CRP without parasitaemia represented acute non-malaria infection, it is plausible that the inverse relationship reflects the development of the immune system with age, which allows older children to be more resilient to infection exposure. Conversely, parasitaemia (with or without fever) was positively associated with age (24-36 months) in the No-iron group. This change in direction of association may suggest that the mechanisms related to infection risk among children in a malaria endemic area could vary across infection types (e.g. malaria vs. non-malaria, or acute vs. chronic). Overall it is surprising that larger differences between the by-group models were not seen, especially considering recent evidence regarding the safety of iron interventions in malaria endemic areas (7, 20, 36). It should be noted, however, that the sample sizes of the by-group sub-analyses (approximately 900 observations) were considerably smaller than those of the combined-group models at endline (n=1780), and thus may not have been powered enough to detect potential associations.

In the combined-group model (**Table 6.4**), higher asset score was associated with a 12% decreased odds of endline infection, defined as CRP >5 mg/L and/or parasitaemia (OR 0.88, 95% CrI 0.78, 0.98). This relationship is not surprising given the well-documented association between household wealth and child health in low- and middle-income settings (37). When compared to our findings at baseline (19), several associations were not consistent over the intervention period; however, the spatial variation in infection risk remained relatively high. This was further illustrated when the predicted means from the combined-group models were

plotted and compared to their respective baseline maps. Plotting the spatial random effect also demonstrated the importance of accounting for residual spatial variation. When infection was defined using CRP only, mapping the residual spatial variation from the combined-group model suggested that the factors included in the analysis may not have fully explained the spatial variation observed. It is possible that a larger sample size and/or geographical coverage may have allowed us to include other covariates from this list and gain additional insight into potential sources of spatial variation; bearing in mind, however, that this may also increase the risk of a Type I error. An additional limitation pertained to CRP and its inability to capture the late phase of the inflammatory response (i.e. after 48 hours), which could have lead to underestimation of non-malaria infection prevalence and thus an incomplete picture of spatial variability.

When the spatial random effect was removed (non-spatial models, **Table 6.5**), there were some changes in the significance of the independent variables, such as elevation and distance to a health facility; however the 95% credible intervals for these associations were often narrow and close to 1.0 (indicating non-significance). These results coincide with other findings in the literature emphasizing the importance of accounting for spatial variation when analyzing geo-referenced data. Hu et al. used Bayesian geostatistical modelling to determine the factors associated with spatial variation in mortality risk among children in China, and found that confidence intervals became wider after taking spatial correlation into account (38). They concluded that ignoring spatial correlation could underestimate the variance of effects, and result in spurious significance of risk factors (38). Ashton et al. used spatial analysis to predict malaria seroprevalence among school children in Ethiopia, and found that including a spatial random effect in their models helped to explain much of the variation between schools (39).

A small compound random effect was observed in all spatial models both at endline and baseline. This suggests that there was low variability in the outcome (infection status) between compounds, regardless of infection definition or intervention group. In rural Ghana, a compound

may consist of more than one household, and thus we expected to find spatial clustering of infection at the compound level. There are two potential explanations for why this was not observed: 1) the compounds within a village were in close proximity to each other, which may have lead to clustering at the village level rather than across compounds; and/or 2) the small number of observations per compound (average cluster size was 1.3) likely reduced the opportunity for the outcome to cluster within compounds, and thus resulted in low variability at the compound level.

Unlike the random effects, the model range increased from baseline to endline across all combined-group spatial models. This change coincided with the appearance of larger and more defined low-risk areas across the study site for all infection definitions (**Figures 6.1-6.4**). Magalhaes et al. described range as the distance at which geographical locations can be considered independent, or in other words, the size of geographical clusters (40). It is possible that the increase in range throughout the study period was driven by the increase in the size of low-risk clusters. The largest low-risk area was focused around the District centre, which is consistent with other studies showing that living in a capital or municipal centre is associated with a reduced risk of adverse infection- or nutrition-related health outcomes (41, 42), and likely due to increased access to health care services.

A potential limitation of the present analyses was the use of straight-line (Euclidean) distance rather than network (travel) distance when estimating proximity to a health facility. Network distance may have been a more appropriate indicator of travel impedance or access since it can more appropriately account for travel by vehicle or bicycle. Calculating distance by road from all study compounds to the nearest health facility or District Centre was not possible due to incomplete or missing vector information (e.g. miss-aligned junctions, missing or disconnected road segments). We considered using the assumption that all travel distances started at the nearest road, though this did not seem appropriate, especially for those villages in more remote locations with highly dispersed compounds. Nesbitt et al. (2014) encountered similar challenges in a study comparing different measures of travel impedance to estimate access to delivery care in the Brong-Ahafo region of Ghana (43). The authors found that straight-line distance was as

informative as network distance for determining spatial access in this setting of rural Ghana [49]. We, therefore, felt it was justified to use Euclidean rather than network distance in the present analysis.

Other limitations include the incomplete representation of all non-malaria infections, and the limited ability to control for ones 'sickliness' or infection susceptibility using only baseline measures. Non-malaria infections were identified using CRP, an acute phase protein that rises in accordance with the early phase of the inflammatory response (approximately 0-2 days). Other acute phase proteins (such as alpha-1-acid glycoprotein) reach their peak concentration during the late phase (approximately 2-8 days), which more closely approximates the change ferritin concentration in response to inflammation (8). Since CRP was the only inflammatory biomarker available for the current analysis, and the study protocol was not designed to include diagnoses of non-malarial infections, it is possible that the prevalence of non-malaria infection among participants was underestimated.

Other than clinical malaria (parasitaemia with fever), the dependent variables used in these analyses were only available at baseline and endline. As such, infection status at baseline was used as an individual-level bio-indicator for general health status and a means of controlling for one's susceptibility to subsequent infections. When it comes to studying infectious disease among children, this analytical practice may be put into question, as individual differences in immune status or antibody production in response to parasitic infections are likely (39) and may be modulated by other infection-specific or sensitive risk factors. To the authors' knowledge, an acceptable solution to this issue does not yet exist, however, the consideration of geographical factors and spatial relationships may help to fill this gap.

Conclusions

The endline analyses presented herein provided an opportunity to explore the spatial dynamics of infection risk in the context of an iron intervention trial in a malaria endemic area among children with varying levels of iron status. The results showed that, in a clinical trial setting where all children received bed nets and access to malaria treatment, there was geographical variation in the risk of infection with distinct low-risk areas, particularly around the District centre. The variation in spatial model outputs observed across infection definitions suggest that it may be important to differentiate between malaria and non-malaria infections when assessing the safety of iron interventions in Ghana or similar malaria endemic settings. Future research should include biomarkers of non-malaria infection that represent both the early and late stages of the acute phase response. Overall, our findings emphasize the importance of considering geographical distribution when assessing infection risk in a paediatric population. They also demonstrate the utility of GIS and spatial analysis methods in communicating population health characteristics that may otherwise not be recognized through traditional regression modelling.

References

1. Children: reducing mortality [Internet]. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs178/en/.

2. WHO. Micronutrient Deficiencies: Iron deficiency anaemia: World Health Organization; 2015 [cited 2012 April]. Available from: http://www.who.int/nutrition/topics/ida/en/.

3. Stoltzfus RJ. Iron deficiency: global prevalence and consequences. Food Nutr Bull. 2003;24(4 Suppl):S99-103. PubMed PMID: 17016951.

4. Katona P, Katona-Apte J. The interaction between nutrition and infection. Clin Infect Dis. 2008;46(10):1582-8. doi: 10.1086/587658. PubMed PMID: 18419494.

5. Spottiswoode N, Duffy PE, Drakesmith H. Iron, anemia and hepcidin in malaria. Front Pharmacol. 2014;5:125. doi: 10.3389/fphar.2014.00125. PubMed PMID: 24910614; PubMed Central PMCID: PMCPMC4039013.

6. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. Science. 2004;306(5704):2090-3. doi: 10.1126/science.1104742. PubMed PMID: 15514116.

7. Sazawal Z, Black, RE., Ramsan, M., Chwaya, HM., Stoltzfus, RJ., Dutta, A ., et al. Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: community-based, randomised, placebo-controlled trial. Lancet. 2006;367:133-43.

8. Thurnham DI, McCabe GP. Influence of infection and inflammation on biomarkers of nutritional status with an emphasis on vitamin A and iron. In: Organization WH, editor. Report: Priorities in the assessment of vitamin A and iron status in populations, Panama City, Panama, 15-17 September 2010. Geneva: World Health Organization; 2012. p. 63-80.

9. Schur N, Vounatsou P, Utzinger J. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. PLoS Negl Trop Dis. 2012;6(9):e1773. doi: 10.1371/journal.pntd.0001773. PubMed PMID: 23029570; PubMed Central PMCID: PMCPMC3441409.

10. Pullan RL, Gething PW, Smith JL, Mwandawiro CS, Sturrock HJ, Gitonga CW, et al. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. PLoS Negl Trop Dis. 2011;5(2):e958. doi: 10.1371/journal.pntd.0000958. PubMed PMID: 21347451; PubMed Central PMCID: PMCPMC3035671.

11. Soares Magalhães RJ, Langa A, Pedro JM, Sousa-Figueiredo JC, Clements AC, Vaz Nery S. Role of malnutrition and parasite infections in the spatial variation in children's anaemia risk in northern Angola. Geospat Health. 2013;7(2):341-54. doi: 10.4081/gh.2013.91. PubMed PMID: 23733295.

12. Root ED, Lucero M, Nohynek H, Anthamatten P, Thomas DS, Tallo V, et al. Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children. Proc Natl Acad Sci U S A. 2014;111(9):3520-5. doi: 10.1073/pnas.1313748111. PubMed PMID: 24550454; PubMed Central PMCID: PMCPMC3948245.

13. Nykiforuk C, Flaman L. Exploring the utilization of geographic information systems in health promotion and public health. Edmonton, AB, Canada: School of Public Health, University of Alberta, 2008.

14. Song P, Zhu Y, Mao X, Li Q, An L. Assessing spatial accessibility to maternity units in Shenzhen, China. PLoS One. 2013;8(7):e70227. doi: 10.1371/journal.pone.0070227. PubMed PMID: 23894622; PubMed Central PMCID: PMCPMC3716609.

15. Masters SH, Burstein R, Amofah G, Abaogye P, Kumar S, Hanlon M. Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis. Soc Sci Med. 2013;93:147-54. doi: 10.1016/j.socscimed.2013.06.012. PubMed PMID: 23906132.

16. Aimone AM, Perumal N, Cole DC. A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: the case of anaemia and malaria. International Journal of Health Geographics. 2013;12:1-13.

17. Musenge E, Vounatsou P, Kahn K. Space-time confounding adjusted determinants of child HIV/TB mortality for large zero-inflated data in rural South Africa. Spat Spatiotemporal Epidemiol. 2011;2(4):205-17. doi: 10.1016/j.sste.2011.07.001. PubMed PMID: 22748220; PubMed Central PMCID: PMCPMC4250009.

18. Samadoulougou S, Maheu-Giroux M, Kirakoya-Samadoulougou F, De Keukeleire M, Castro MC, Robert A. Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. Parasit Vectors. 2014;7:350. doi: 10.1186/1756-3305-7-350. PubMed PMID: 25074132; PubMed Central PMCID: PMCPMC4262087.

19. Aimone Phillips AM, Brown PE, Owusu-Agyei S, Zlotkin SH, Cole DC. Geographical patterns and associated risks of infection among Ghanaian children: a secondary spatial analysis. 2015.

20. Zlotkin S, Newton S, Aimone AM, Azindow I, Amenga-Etego S, Tchum K, et al. Effect of iron fortification on malaria incidence in infants and young children in Ghana: a randomized trial. JAMA. 2013;310(9):938-47. doi: 10.1001/jama.2013.277129. PubMed PMID: 24002280.

21. WHO. World Malaria Report 2009. Geneva, Switzerland: World Health Organization, 2009.

22. WHO. Worldwide prevalence of anaemia 1993-2005. Geneva, Switzerland: World Health Organization, 2008.

23. WHO/UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children: A Joint Statement by the World Health Organization and the United Nations Children's Fund. Geneva, Switzerland: World Health Organization, 2009 Contract No.: ISBN 978 92 4 159816 3.

24. USGS. United States Geological Survey 2012 [cited 2013 January]. Available from: https://lta.cr.usgs.gov.

25. USGS. Land Processes Distributed Active Archive Center 2014 [cited 2013 January]. Available from: https://lpdaac.usgs.gov.

26. Hengl T. Worldgrids — a public repository and a WPS for global environmental layers: ISRIC - World Soil Information; 2014 [cited 2013 January]. Available from: http://worldgrids.org.

27. Adish AA, Esrey SA, Gyorkos TW, Johns T. Risk factors for iron deficiency anaemia in preschool children in northern Ethiopia. Public Health Nutrition. 1999;2(3):243-52. PubMed PMID: 1999331372.

28. Jonker FA, Calis JC, van Hensbroek MB, Phiri K, Geskus RB, Brabin BJ, et al. Iron status predicts malaria risk in Malawian preschool children. PLoS One. 2012;7(8):e42670. doi: 10.1371/journal.pone.0042670. PubMed PMID: 22916146; PubMed Central PMCID: PMCPMC3420896.

29. Namaste S, Rohner R, Suchdev P, Kupka R, Mei Z, Bhushan N, et al. Approaches to address the association between inflammation and ferritin in children and women. Unpublished.

30. Diggle PJ, Moyeed RA, Tawn JA. Model-based geostatistics. Applied Statistics. 1998;47:299-350.

31. Diggle PJ, Ribeiro PJ. Model-based geostatistics. New York: Springer-Verlag; 2006.

32. Rue H, Martino S. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B. 2009;71(2):319-92.

33. Lindgren F, Rue H, Lindstrom J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society (Series B, Statistical Methodology). 2011;73, Part 4:423-98.

34. Brown PE. Model-based geostatistics the easy way. Journal of Statistical Software. 2015;63(12):1-24.

35. R. R: A Language and Environment for Statistical Computing. 3.2.2 ed. Vienna, Austria: R Foundation for Statistical Computing; 2015.

36. Jaeggi T, Kortman GA, Moretti D, Chassard C, Holding P, Dostal A, et al. Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. Gut. 2015;64(5):731-42. doi: 10.1136/gutjnl-2014-307720. PubMed PMID: 25143342.

37. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet. 2008;371(9608):243-60. doi: S0140-6736(07)61690-0 [pii]

10.1016/S0140-6736(07)61690-0. PubMed PMID: 18207566.

38. Hu Y, Wang JF, Li XH, Ren D, Mu DH, Wang YP, et al. Application of Bayesian geostatistical modeling for the assessment of risk for child mortality during the 2008 earthquake in Wenchuan, People's Republic of China. Geospat Health. 2012;6(2):247-55. doi: 10.4081/gh.2012.142. PubMed PMID: 22639126.

39. Ashton RA, Kefyalew T, Rand A, Sime H, Assefa A, Mekasha A, et al. Geostatistical modeling of malaria endemicity using serological indicators of exposure collected through school surveys. Am J Trop Med Hyg. 2015;93(1):168-77. doi: 10.4269/ajtmh.14-0620. PubMed PMID: 25962770; PubMed Central PMCID: PMCPMC4497890.

40. Soares Magalhães RJ, Clements AC. Spatial heterogeneity of haemoglobin concentration in preschool-age children in sub-Saharan Africa. Bull World Health Organ. 2011;89(6):459-

68. doi: 10.2471/BLT.10.083568. PubMed PMID: 21673862; PubMed Central PMCID: PMCPMC3099553.

41. Shirayama Y, Phompida S, Shibuya K. Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. Malar J. 2009;8:217. doi: 10.1186/1475-2875-8-217. PubMed PMID: 19772628; PubMed Central PMCID: PMCPMC2754997.

42. Kandala NB, Madungu TP, Emina JB, Nzita KP, Cappuccio FP. Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? BMC Public Health. 2011;11:261. doi: 10.1186/1471-2458-11-261. PubMed PMID: 21518428; PubMed Central PMCID: PMCPMC3111378.

43. Nesbitt RC, Gabrysch S, Laub A, Soremekun S, Manu A, Kirkwood BR, et al. Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. Int J Health Geogr. 2014;13:25. doi: 10.1186/1476-072X-13-25. PubMed PMID: 24964931; PubMed Central PMCID: PMCPMC4086697.

Table 6.1: Baseline and endline characteristics of the Ghana trial participants

	Baseline	Endline
Trial participants with a geo-coded residence (n)	1943	1780
Males (%)	992 (51.1)	900 (50.5)
Age at enrolment (months), mean (SD)[a]	19.2 (8.5)	19.3 (8.5)
Serum ferritin (µg/L), geometric mean (SD)	35.1 (3.65)	73.9 (3.58)
Infection status		
C-reactive protein (mg/L), mean (SD)	3.34 (4.96)	3.86 (5.10)
Parasite density (count/µL), geometric mean (SD)	3003.0 (5.35)	4160.4 (6.30)
Inflammation and/or parasitaemia[b], n (%)	719 (37.0)	741 (41.6)
Inflammation without parasitaemia[c], n (%)	272 (14.0)	258 (14.5)
Parasitaemia with fever[d][f], n (%)	150 (7.72)	555 (28.6)
Parasitaemia[e], n (%)	447 (23.0)	483 (27.1)
Anthropometric status[f]		
Weight-for-length z-score, mean (SD)	-0.63 (0.97)	-0.62 (0.97)
Length-for-age z-score, mean (SD)	-0.81 (1.21)	-0.80 (1.29)
Maternal Education, n (%)[g]		
None	586 (33.5)	543 (33.0)
Any	1166 (66.5)	1105 (67.0)
Household Asset Score, n (%)[h]		
Low	900 (49.0)	817 (47.7)
High	938 (51.0)	897 (52.3)

[f]Endline value represents the period prevalence for 1939 participants

[a]Measured at baseline only

[b]Inflammation and/or parasitaemia (n=1780) = CRP > 5 mg/L and/or any malaria parasitaemia;

[c]Inflammation without parasitaemia (n=1780) = CRP > 5 mg/L without malaria parasitaemia;

[d]Parasitaemia with fever (n=1939) = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

[e]Parasitaemia (n=1780) = any malaria parasitaemia (with/without fever)

[f]Measured at baseline; z-scores estimated using the WHO Child Growth Standards [23]

[g]Measured at baseline only; total n=1752 (74 respondents were not mothers, 117 missing due to incomplete surveys)

[h] Measured at baseline only; reduced sample size (approximately 1825) due to incomplete surveys and "unknown" responses

Table 6.2: Results from the final model of geospatial and non-geospatial risk factors of endline infection status[1] among Ghanaian children in the *No-iron* group (Brong-Ahafo Region, May-November 2010)

Covariates	Estimate (95% CrI) on the exponential scale[2]		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia		**(n=894)**			
Intercept	0.550	(0.298, 1.034)	7.050	0.506	0.008
Age per month			(2.016, 18.26)	(0.271, 0.979)	(0.004, 0.030)
6-23 months	1.004	(0.975, 1.035)			
24-35 months	1.018	(0.964, 1.074)			
Sex (male reference)	1.016	(0.760, 1.357)			
Length-for-age z-score	0.984	(0.869, 1.113)			
Weight-for-length z-score	1.229	(1.056, 1.431)			
Asset score	0.893	(0.760, 1.047)			
Distance to health facility (km)	1.072	(0.963, 1.188)			
Elevation (m)	0.995	(0.990, 1.001)			
Baseline infection status	1.887	(1.384, 2.576)			
Baseline iron status	1.051	(0.938, 1.179)			

2) Inflammation without parasitaemia		**(n=894)**			
Intercept	0.219	(0.109, 0.426)	9.458	0.442	0.007
Age per month			(2.389, 23.57)	(0.206, 0.965)	(0.004, 0.029)
6-23 months	1.011	(0.971, 1.052)			
24-35 months	0.920	(0.843, 0.996)			
Sex (male reference)	1.001	(0.671, 1.493)			
Length-for-age z-score	1.153	(0.977, 1.360)			
Weight-for-length z-score	1.302	(1.058, 1.603)			
Asset score	0.872	(0.705, 1.075)			
Distance to health facility (km)	1.065	(0.953, 1.176)			
Elevation (m)	0.997	(0.991, 1.003)			
Baseline infection status	1.034	(0.556, 1.813)			
Baseline iron status	1.003	(0.841, 1.166)			
3) Parasitaemia with fever (n=979)					
Intercept	0.003	(0.002, 0.004)	7.690	0.448	0.007
Age per month			(2.821, 18.01)	(0.243, 0.858)	(0.004, 0.028)
6-23 months	1.001	(0.979, 1.023)			
24-35 months	0.970	(0.927, 1.013)			
Sex (male reference)	0.896	(0.718, 1.116)			
Length-for-age z-score	1.009	(0.919, 1.107)			
Weight-for-length z-score	0.896	(0.797, 1.007)			

Asset score	1.075	(0.947, 1.220)			
Distance to health facility (km)	1.047	(0.959, 1.142)			
Elevation (m)	0.998	(0.993, 1.003)			
Baseline infection status	0.757	(0.467, 1.161)			
Baseline iron status	1.040	(0.953, 1.125)			
4) Parasitaemia (n=894)					
Intercept	0.186	(0.089, 0.400)	6.535	0.652	0.008
Age per month			(2.037, 15.08)	(0.344, 1.255)	(0.004, 0.029)
6-23 months	0.974	(0.936, 1.014)			
24-35 months	1.067	(1.026, 1.110)			
Sex (male reference)	1.050	(0.756, 1.459)			
Length-for-age z-score	0.901	(0.780, 1.037)			
Weight-for-length z-score	1.095	(0.923, 1.298)			
Asset score	0.941	(0.781, 1.130)			
Distance to health facility (km)	1.023	(0.896, 1.165)			
Elevation (m)	0.996	(0.989, 1.004)			
Baseline infection status	2.864	(1.968, 4.172)			
Baseline iron status	1.036	(0.914, 1.169)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (count): any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

[2]Exponentials of posterior medians and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases the risk of infection by 5% (odds ratio from logistic model for all binary outcomes; relative risk from Poisson model for count outcome)

CrI = credible interval

Baseline infection status = baseline infection status (yes/no) according to corresponding definition

Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range

Table 6.3: Results from the final model of geospatial and non-geospatial risk factors of endline infection status[1] among Ghanaian children in the *Iron* group (Brong-Ahafo Region, May-November 2010)

Covariates	Estimate (95% CrI) on the exponential scale[2]		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia	**(n=886)**				
Intercept	0.647	(0.354, 1.263)	7.761	0.487	0.007
Age per month			(2.682, 17.96)	(0.263, 0.931)	(0.004, 0.028)
6-23 months	1.004	(0.975, 1.035)			
24-35 months	0.977	(0.925, 1.031)			
Sex (male reference)	1.162	(0.871, 1.551)			
Length-for-age z-score	0.980	(0.857, 1.119)			
Weight-for-length z-score	0.971	(0.833, 1.130)			
Asset score	0.906	(0.769, 1.064)			
Distance to health facility (km)	1.003	(0.904, 1.106)			
Elevation (m)	0.995	(0.989, 1.001)			
Baseline infection status	2.272	(1.664, 3.108)			
Baseline iron status	1.053	(0.953, 1.172)			

2) Inflammation without parasitaemia		**(n=886)**			
Intercept	0.143	(0.073, 0.274)	12.56 (4.318,	0.428 (0.203,	0.007 (0.004,
Age per month			27.27)	0.923)	0.028)
6-23 months	1.007	(0.969, 1.048)			
24-35 months	0.983	(0.910, 1.056)			
Sex (male reference)	1.330	(0.905, 1.962)			
Length-for-age z-score	1.054	(0.884, 1.255)			
Weight-for-length z-score	0.890	(0.726, 1.090)			
Asset score	0.888	(0.725, 1.087)			
Distance to health facility (km)	0.984	(0.883, 1.080)			
Elevation (m)	1.001	(0.995, 1.006)			
Baseline infection status	1.508	(0.897, 2.459)			
Baseline iron status	1.044	(0.918, 1.166)			
3) Parasitaemia with fever (n=960)					
Intercept	0.002	(0.001, 0.004)	8.352 (3.203,	0.466 (0.244,	0.007 (0.004,
Age per month			17.99)	0.896)	0.028)
6-23 months	1.017	(0.992, 1.043)			
24-35 months	0.987	(0.945, 1.029)			
Sex (male reference)	0.991	(0.780, 1.257)			
Length-for-age z-score	0.969	(0.867, 1.082)			
Weight-for-length z-score	1.052	(0.925, 1.197)			

Asset score	1.002	(0.872, 1.152)			
Distance to health facility (km)	1.064	(0.970, 1.163)			
Elevation (m)	0.998	(0.993, 1.003)			
Baseline infection status	0.849	(0.516, 1.321)			
Baseline iron status	0.981	(0.895, 1.060)			
4) Parasitaemia (n=886)					
Intercept	0.412	(0.199, 0.938)	7.293 (2.486,	0.666 (0.376,	0.008 (0.004,
Age per month			17.67)	1.230)	0.030)
6-23 months	1.015	(0.975, 1.056)			
24-35 months	0.977	(0.938, 1.016)			
Sex (male reference)	0.977	(0.707, 1.347)			
Length-for-age z-score	0.946	(0.812, 1.098)			
Weight-for-length z-score	1.034	(0.872, 1.225)			
Asset score	0.946	(0.785, 1.137)			
Distance to health facility (km)	1.009	(0.889, 1.149)			
Elevation (m)	0.994	(0.987, 1.001)			
Baseline infection status	2.597	(1.791, 3.768)			
Baseline iron status	1.042	(0.939, 1.153)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (count): any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

[2]Exponentials of posterior medians and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases the risk of infection by 5% (odds ratio from logistic model for all binary outcomes; relative risk from Poisson model for count outcome)

CrI = credible interval

Baseline infection status = baseline infection status (yes/no) according to corresponding definition

Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range

Table 6.4: Results from the final model of geospatial and non-geospatial risk factors of endline infection status[1] among Ghanaian children in both the *Iron* and *No-iron* groups (Brong-Ahafo Region, May-November 2010)

Covariates	Estimate (95% CrI) on the exponential scale[2]		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia (n=1780)					
Intercept	0.589	(0.337, 1.072)	7.341	0.525	0.008
Age per month			(2.928, 15.84)	(0.306, 0.944)	(0.004, 0.029)
6-23 months	1.004	(0.984, 1.026)			
24-35 months	1.000	(0.962, 1.038)			
Sex (male reference)	1.087	(0.886, 1.333)			
Length-for-age z-score	0.982	(0.896, 1.075)			
Weight-for-length z-score	1.088	(0.978, 1.211)			
Asset score	0.875	(0.779, 0.980)			
Distance to health facility (km)	1.034	(0.933, 1.136)			
Elevation (m)	0.996	(0.991, 1.001)			
Baseline infection status	1.843	(1.359, 2.500)			
Group	1.094	(0.826, 1.448)			

Baseline iron status	1.068	(0.956, 1.194)			
Baseline infection status*Group	1.192	(0.775, 1.833)			
Baseline iron status*Group	0.983	(0.847, 1.144)			
2) Inflammation without parasitaemia (n=1780)					
Intercept	0.183	(0.106, 0.309)	10.77	0.407	0.007
Age per month			(8.767, 25.54)	(0.208, 0.830)	(0.004, 0.028)
6-23 months	1.009	(0.982, 1.038)			
24-35 months	0.956	(0.904, 1.009)			
Sex (male reference)	1.139	(0.865, 1.502)			
Length-for-age z-score	1.092	(0.968, 1.230)			
Weight-for-length z-score	1.054	(0.912, 1.217)			
Asset score	0.871	(0.750, 1.010)			
Distance to health facility (km)	1.012	(0.924, 1.093)			
Elevation (m)	0.999	(0.994, 1.003)			
Baseline infection status	1.055	(0.574, 1.826)			
Group	1.009	(0.713, 1.427)			
Baseline iron status	1.001	(0.844, 1.156)			
Baseline infection status*Group	1.523	(0.718, 3.300)			
Baseline iron status*Group	1.033	(0.856, 1.261)			
3) Parasitaemia with fever (n=1939)					
Intercept	0.003	(0.002, 0.004)	7.365	0.464	0.007

Age per month			(3.225, 15.18)	(0.273, 0.825)	(0.004, 0.028)
6-23 months	1.008	(0.991, 1.025)			
24-35 months	0.978	(0.948, 1.008)			
Sex (male reference)	0.939	(0.799, 1.104)			
Length-for-age z-score	0.992	(0.923, 1.064)			
Weight-for-length z-score	0.966	(0.886, 1.053)			
Asset score	1.029	(0.937, 1.130)			
Distance to health facility (km)	1.057	(0.972, 1.149)			
Elevation (m)	0.997	(0.993, 1.002)			
Group	0.891	(0.729, 1.087)			
Baseline infection status	0.730	(0.452, 1.114)			
Baseline iron status	1.028	(0.942, 1.112)			
Baseline infection status*Group	1.209	(0.639, 2.282)			
Baseline iron status*Group	0.954	(0.849, 1.069)			
4) Parasitaemia (n=1780)					
Intercept	0.271	(0.135, 0.571)	7.846	0.712	0.008
Age per month			(3.384, 16.06)	(0.406, 1.291)	(0.004, 0.029)
6-23 months	0.993	(0.966, 1.021)			
24-35 months	1.022	(0.994, 1.050)			
Sex (male reference)	1.024	(0.814, 1.287)			
Length-for-age z-score	0.926	(0.835, 1.026)			

Weight-for-length z-score	1.068	(0.947, 1.204)
Asset score	0.916	(0.802, 1.044)
Distance to health facility (km)	1.022	(0.902, 1.164)
Elevation (m)	0.996	(0.990, 1.003)
Baseline infection status	2.746	(1.910, 3.951)
Group	1.169	(0.862, 1.587)
Baseline iron status	1.061	(0.940, 1.193)
Baseline infection status*Group	0.915	(0.555, 1.508)
Baseline iron status*Group	0.980	(0.839, 1.145)

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (count): any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

[2]Exponentials of posterior medians and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases the risk of infection by 5% (odds ratio from logistic model for all binary outcomes; relative risk from Poisson model for count outcome)

CrI = credible interval

Baseline infection status = baseline infection status (yes/no) according to corresponding definition

Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range

Table 6.5: Results from the final combined-group spatial and non-spatial models of risk factors of endline infection status[1] among Ghanaian children (Brong-Ahafo Region, May-November 2010)

	Total Spatial		Total Non-Spatial	
Covariates	**Estimate (95% CrI) on the exponential scale[2]**		**Estimate (95% CrI) on the exponential scale[2]**	
Inflammation and/or parasitaemia	**(n=1780)**			
Intercept	0.589	(0.337, 1.072)	0.478	**(0.344, 0.660)**
Age per month				
6-23 months	1.004	(0.984, 1.026)	1.002	(0.981, 1.023)
24-35 months	1.000	(0.962, 1.038)	1.001	(0.964, 1.039)
Sex (male reference)	1.087	(0.886, 1.333)	1.069	(0.875, 1.307)
Length-for-age z-score	0.982	(0.896, 1.075)	0.965	(0.883, 1.055)
Weight-for-length z-score	1.088	(0.978, 1.211)	1.065	(0.959, 1.183)
Asset score	0.875	(0.779, 0.980)	0.991	(0.895, 1.098)
Distance to health facility (km)	1.034	(0.933, 1.136)	1.043	**(1.006, 1.081)**
Elevation (m)	0.996	(0.991, 1.001)	0.995	**(0.993, 0.997)**
Baseline infection status	1.843	(1.359, 2.500)	1.882	**(1.396, 2.539)**
Group	1.094	(0.826, 1.448)	1.082	(0.821, 1.427)
Baseline iron status	1.001	(0.999, 1.002)	1.064	(0.954, 1.188)
Baseline infection status*Group	1.192	(0.775, 1.833)	1.183	(0.774, 1.808)
Baseline iron status*Group	1.000	(0.998, 1.002)	0.986	(0.851, 1.146)

Inflammation without parasitaemia		**(n=1780)**		
Intercept	0.183	(0.106, 0.309)	0.178	**(0.115, 0.272)**
Age per month				
6-23 months	1.009	(0.982, 1.038)	1.010	(0.982, 1.038)
24-35 months	0.956	(0.904, 1.009)	0.955	(0.903, 1.007)
Sex (male reference)	1.139	(0.865, 1.502)	1.140	(0.866, 1.502)
Length-for-age z-score	1.092	(0.968, 1.230)	1.086	(0.963, 1.224)
Weight-for-length z-score	1.054	(0.912, 1.217)	1.048	(0.907, 1.209)
Asset score	0.871	(0.750, 1.010)	0.916	(0.798, 1.053)
Distance to health facility (km)	1.012	(0.924, 1.093)	1.030	(0.982, 1.081)
Elevation (m)	0.999	(0.994, 1.003)	0.998	(0.995, 1.002)
Baseline infection status	1.055	(0.574, 1.826)	1.017	(0.555, 1.754)
Group	1.009	(0.713, 1.427)	0.984	(0.696, 1.391)
Baseline iron status	1.000	(0.998, 1.002)	0.991	(0.836, 1.145)
Baseline infection status*Group	1.523	(0.718, 3.300)	1.549	(0.732, 3.347)
Baseline iron status*Group	1.000	(0.998, 1.003)	1.045	(0.865, 1.276)
Parasitaemia with fever (n=1939)				
Intercept	0.003	(0.002, 0.004)	0.002	**(0.002, 0.003)**
Age per month				
6-23 months	1.008	(0.991, 1.025)	1.007	(0.991, 1.024)
24-35 months	0.978	(0.948, 1.008)	0.978	(0.948, 1.007)

Sex (male reference)	0.939	(0.799, 1.104)	0.940	(0.800, 1.103)
Length-for-age z-score	0.992	(0.923, 1.064)	0.981	(0.914, 1.052)
Weight-for-length z-score	0.966	(0.886, 1.053)	0.951	(0.873, 1.035)
Asset score	1.029	(0.937, 1.130)	1.130	**(1.039, 1.230)**
Distance to health facility (km)	1.057	(0.972, 1.149)	1.057	**(1.030, 1.085)**
Elevation (m)	0.997	(0.993, 1.002)	0.999	(0.997, 1.001)
Group	0.891	(0.729, 1.087)	0.900	(0.738, 1.096)
Baseline infection status	0.730	(0.452, 1.114)	0.741	(0.460, 1.128)
Baseline iron status	1.000	(0.999, 1.001)	1.039	(0.955, 1.120)
Baseline infection status*Group	1.209	(0.639, 2.282)	1.208	(0.641, 2.271)
Baseline iron status*Group	0.999	(0.998, 1.001)	0.949	(0.846, 1.060)
Parasitaemia (n=1780)				
Intercept	0.271	(0.135, 0.571)	0.207	**(0.146, 0.291)**
Age per month				
6-23 months	0.993	(0.966, 1.021)	0.987	(0.961, 1.015)
24-35 months	1.022	(0.994, 1.050)	1.024	(0.997, 1.052)
Sex (male reference)	1.024	(0.814, 1.287)	1.005	(0.803, 1.257)
Length-for-age z-score	0.926	(0.835, 1.026)	0.915	(0.827, 1.010)
Weight-for-length z-score	1.068	(0.947, 1.204)	1.037	(0.922, 1.166)
Asset score	0.916	(0.802, 1.044)	1.034	(0.920, 1.162)
Distance to health facility (km)	1.022	(0.902, 1.164)	1.021	(0.981, 1.061)

Elevation (m)	0.996	(0.990, 1.003)	0.995	**(0.993, 0.998)**
Baseline infection status	2.746	(1.910, 3.951)	2.833	(**1.992, 4.030)**
Group	1.169	(0.862, 1.587)	1.167	(0.865, 1.574)
Baseline iron status	1.001	(0.999, 1.002)	1.069	(0.951, 1.198)
Baseline infection status*Group	0.915	(0.555, 1.508)	0.940	(0.576, 1.533)
Baseline iron status*Group	1.000	(0.998, 1.002)	0.972	(0.835, 1.133)

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (count): any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours);

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

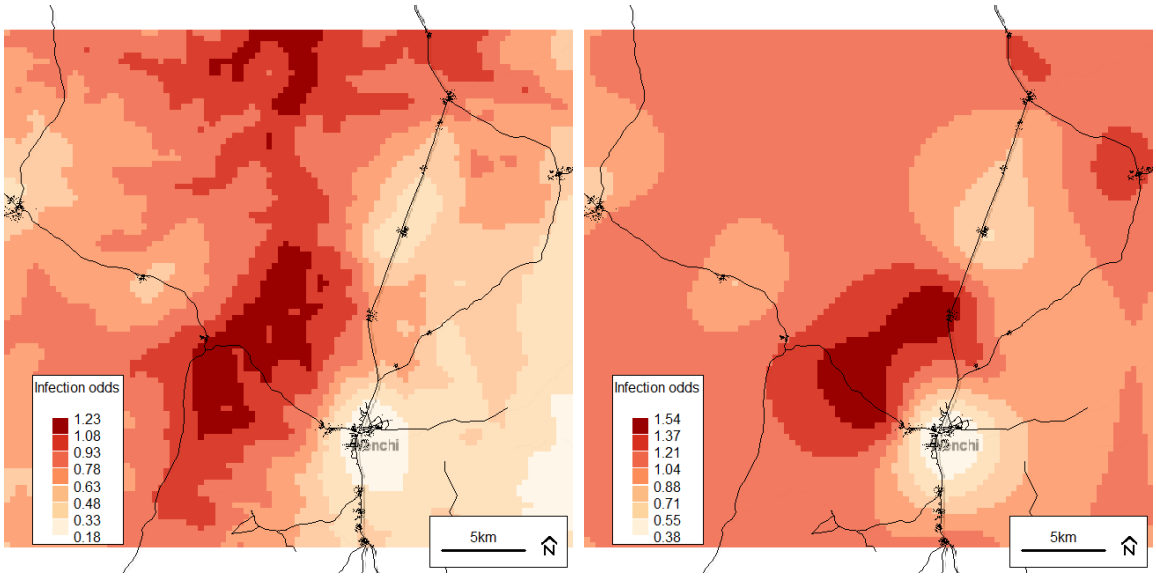
Model prior shape = 1.117, model prior rate = 0.157

[2]Exponentials of posterior medians and 2.5% and 97.5% posterior quantiles of model parameters, with a value of 1.05 indicating that a variable increases the risk of infection by 5% (odds ratio from logistic model for all binary outcomes; relative risk from Poisson model for count outcome)

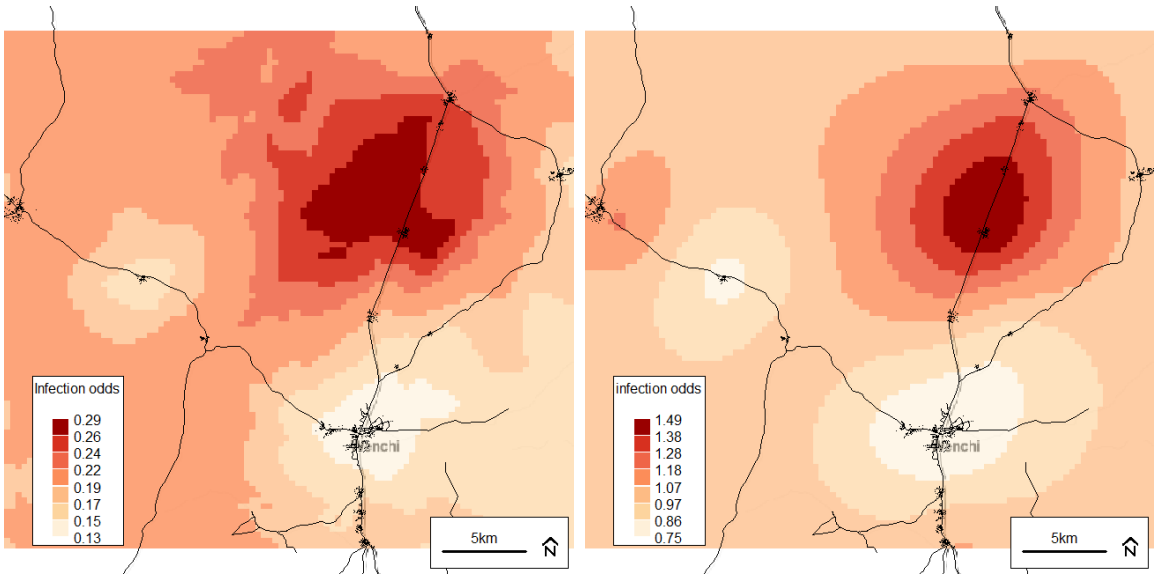
CrI = credible interval

Baseline infection status = baseline infection status (yes/no) according to corresponding definition

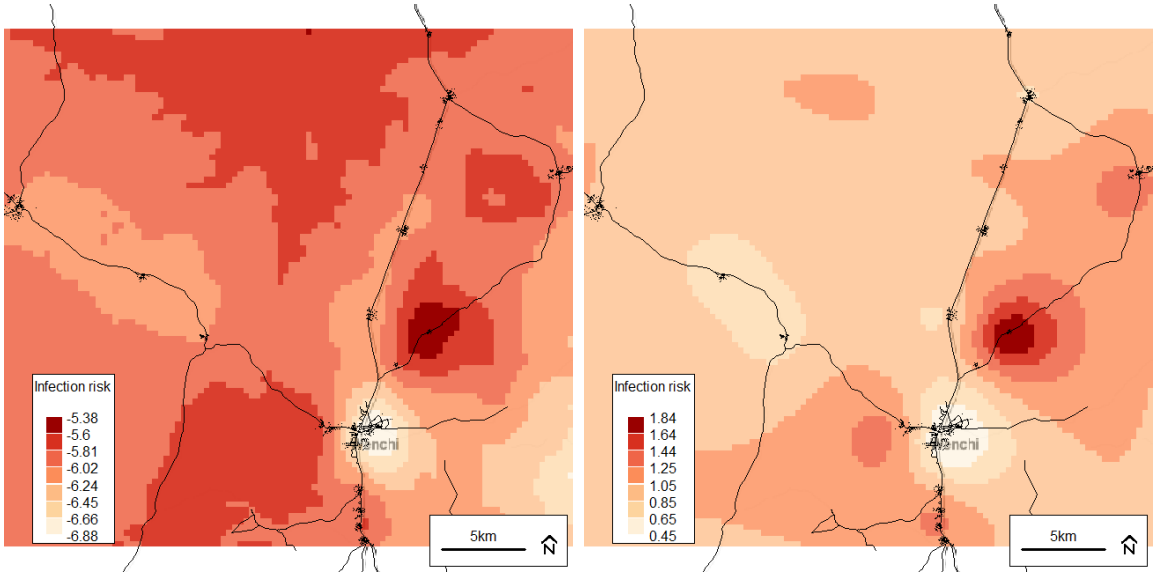
Baseline iron status = iron status at baseline, defined as serum ferritin concentration (µg/dL) corrected for CRP using the regression method (Namaste et al.) and re-scaled by multiplying by the inverse of the inter-quartile range



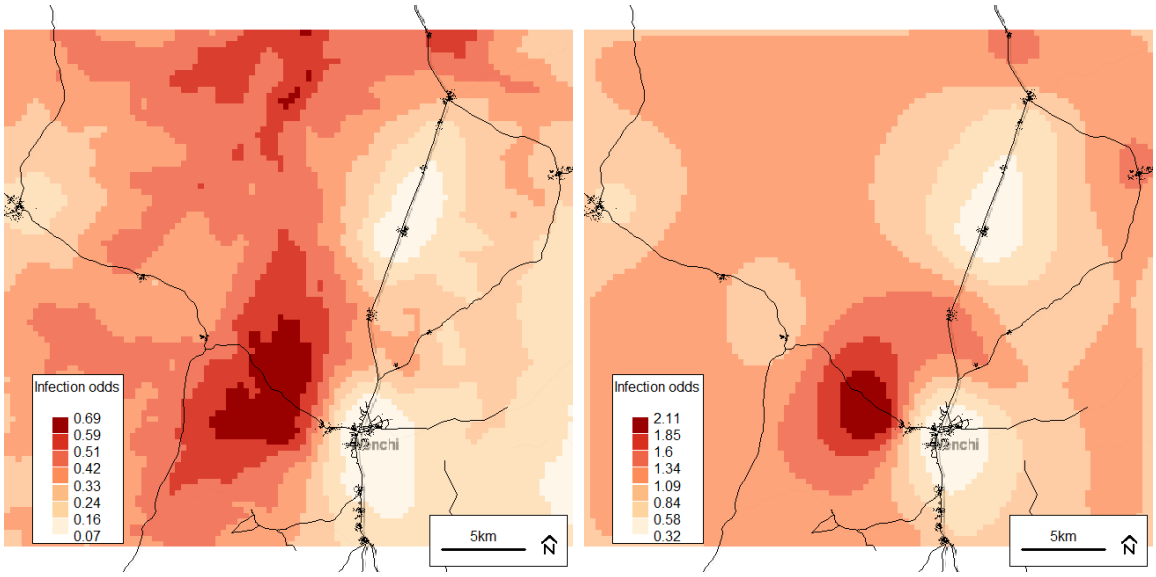
Figures 6.1a and 6.1b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final combined-group model (Model 4). Darker colour indicates higher risk (odds) of inflammation (CRP>5 mg/L) *and/or any* malaria parasitaemia at endline. Background © Stamen Design.



Figures 6.2a and 6.2b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final combined-group model (Model 4). Darker colour indicates higher risk (odds) of inflammation (CRP > 5 mg/L) *without* malaria parasitaemia at endline. Background © Stamen Design.



Figures 6.3a and 6.3b: Predicted median raster plot (left), and residual spatial variation plot (right) from the final combined-group model (Model 4). Darker colour indicates higher risk of malaria parasitaemia *with* concurrent fever (axillary temperature >37.5^{0}C - or history of reported fever within 48 hours)at endline. Background © Stamen Design.



Figures 6.4a and 6.4b: Predicted odds raster plot (left), and residual spatial variation plot (right) from the final combined-group model (Model 4). Darker colour indicates higher risk (odds) of malaria parasitaemia *with or without* fever at endline. Background © Stamen Design.

Chapter 7
Discussion

7 Discussion

7.1 Summary of thesis findings in relation to specific aims

The geostatistical analyses conducted herein have utilized Bayesian inference with integrated nested Laplace approximations (INLA) to explore the spatial variation and associated risk factors of iron status and infection among young Ghanaian children before and after a 5-month randomized iron intervention trial.

In Aim 1, I sought to determine whether geo-indicators could be used to assess the risk (or risk potential) of iron deficiency among the study participants both at baseline and endline. The findings demonstrated the challenge of assessing iron status using serum ferritin in children who are at high risk of infection exposure, particularly malaria. Although the ferritin values were corrected for inflammation (CRP), the lack of associations observed with or without iron treatment highlights the challenges of assessing iron status in a malaria endemic area, as well as the complexity of the mechanisms behind maintaining iron homeostasis in the presence of inflammation and infection. I addressed this complexity by exploring infection as a dependent variable using similar spatial modelling methods (Aim 2).

In Aim 2, the findings were more promising in terms of identifying geo-spatial factors associated with baseline infection status; however, data limitations continued to present challenges with respect to defining non-malaria infections. In an attempt to address these data limitations, four different infection dependent variables were created using combinations of CRP concentration, parasite count, and fever. Considering the variation in model outputs across infection definitions, it seemed that parasitaemia was the main driver behind the spatial

relationships observed, though the cross sectional nature of the baseline data did not allow causal inferences to be made.

In order to further explore the baseline relationships, I extended the baseline spatial models to include endline infection status as the dependent variable (Aim 3). Similar to the baseline analyses, four different endline infection dependent variables were created using combinations of CRP concentration, parasite count, and fever. The endline analyses were also stratified by intervention group to determine if the effects of time and treatment had differential impacts on the infection outcomes. Overall, the by-group and combined-group analyses showed that the effect of time and iron treatment seemed to diminish the relationships observed at baseline; however, comparisons to a non-spatial model indicated that including a spatial random effect may have also contributed to the weakened associations.

In fact, although several associations found at baseline were not observed at the end of the intervention period for both the iron and infection outcomes, the spatial variation remained high. This suggests that geographical distribution may be an important feature that could help to improve our understanding of iron and infection dynamics in malaria endemic areas. For example, comparing the plots of the predicted mean and spatial random effects for iron status and infection status revealed consistent similarities in spatial variation both at baseline and endline. In other words, areas where ferritin concentrations were lower tended to coincide with those of low infection risk (defined using parasitaemia or CRP). Despite our efforts to account for potential interactions between iron status and infection risk (i.e. correcting ferritin for inflammation and including iron status as a covariate in infection models), the similarities in the spatial distribution of each may have two possible explanations: 1) Correcting ferritin for the effect of infection using CRP only was not sufficient, and thus the distribution of infection was a large contributor to variations in iron status; or 2) Those with lower iron status had a lower risk of infection due to the inhibitive effect of iron-restriction on pathogen growth. The former explanation is plausible, since inflammation (associated with infection) has an up-regulating effect on acute phase proteins like ferritin (D.I. Thurnham & McCabe, 2012). Therefore, if ferritin correction was successful, then I might have expected to see spatial variation in iron

status at endline that was more influenced by the geographical distribution of those in Iron vs. No-iron intervention groups rather than the variation in infection status per se.

The latter explanation reflects a well-documented relationship between infection and iron homeostasis that is particularly evident in settings where parasitic infections are common (Brabin, Brabin, & Gies, 2013; M. A. Clark, Goheen, & Cerami, 2014). If it were true (i.e. those with lower iron status had a lower risk of infection), then I would expect to find lower infection risk at endline among those in the Iron group who were iron deficient. Indeed the by-group analyses suggest this (odds ratio for baseline iron status ranged from 1.02 to 1.03); however, baseline iron status was also positively (though not significantly) associated with endline infection status in the No-iron group and both analyses demonstrated strong positive associations with baseline infection status in models of infection defined using parasitaemia (with or without high CRP or fever). Further, there was a higher overall prevalence of infection at endline, which appeared to reflect an increase in parasitaemia, though was also accompanied by a higher prevalence of inflammation (CRP >5 mg/L) among children in the Iron group. These changes in infection biomarkers also coincided with an increase in ferritin concentration from baseline to endline, which equated to only a minor difference in iron deficiency prevalence between interventions groups (approximately 4% in the Iron group and 5% in the No-iron group). This finding is surprising given the magnitude of effect of iron-containing MNPs reported in other studies of young children (De-Regil, Suchdev, Vist, Walleser, & Pena-Rosas, 2011; P. S. Suchdev et al., 2012). It has been found that the normal immune response to infection and inflammation involves the up-regulation of hepcidin (an iron regulatory hormone), which prevents the passage of iron into the plasma and reduces iron absorption, thus inhibiting further microbial growth (R. F. Hurrell, 2011). Considering that the Ghana trial was conducted during the rainy season (when malaria transmission tends to be higher), it is possible that the efficacy of the iron-MNPs was reduced due to an elevated exposure to infection during the intervention period. Therefore, the spatial patterns and associated risks of iron status observed at baseline and endline may have been primarily driven by malaria and non-malaria infection status rather than the intervention itself.

Had the intervention demonstrated larger differential effects across groups, it is possible that I would have observed more significant spatial variable associations at endline. However, another modifying factor to consider is the effect of other infection control activities during the intervention period. In the Ghana trial, although only half of the participants received MNPs with iron, they were all provided with malaria treatment when indicated, insecticide-treated nets, and improved access to the health system. Both intervention groups also received additional micronutrients that may have had a general enhancing effect on immune function (e.g. vitamin A and zinc). All (or a combination of) these factors may have overridden the efficacy of the iron intervention, and thus reduced the strength of potential spatial predictors.

7.2 Thesis contributions to the Iron and Malaria literature

While the findings from each Aim may not provide clarity in terms of the safety of iron interventions in malaria endemic areas, they do contribute to the "Iron and Malaria" literature by adding a geographical dimension to our current understanding of iron and infection risk dynamics. In terms of the research priorities of the "Iron and Malaria Project", the analyses presented here contribute to all three pillars, which include mechanisms, biomarkers, and interventions.

Most of the recent evidence generated to date on potential *mechanisms* that explain the adverse effects of iron supplementation on infection risk has come from the basic sciences. For example, Clark et al. have used *in vitro* experimentation to show that the labile iron pool (LIP) of parasitized erythrocytes increases as the malaria parasite develops and with the addition of either transferrin or ferric citrate to culture media, indicating that the parasite is able to access host serum iron sources (M. Clark, Fisher, Kasthuri, & Cerami Hand, 2013). Clark et al. have also demonstrated that *Plasmodium falciparum* infects iron deficient erythrocytes less efficiently, and that iron supplementation of iron deficient individuals could reverse the protective effects of iron deficiency, likely due to merozoite preference for young erythrocytes, which are more abundant during recovery from iron deficiency anaemia due to transient

reticulocytosis (M. A. Clark, Goheen, Fulford, et al., 2014). Overall these findings suggest that malaria parasites are able to access host serum iron sources (M. Clark et al., 2013), and also provide experimental validation of field observations reporting protective effects of iron deficiency and harmful effects of iron administration on human malaria susceptibility (M. A. Clark, Goheen, Fulford, et al., 2014). Hod et al. used double, leukoreduced, RBC units from healthy human volunteers to determine the differential impacts of transfusion with autologous fresh RBC units (after storage for 3-7 days) versus older units (after storage for 40-42 days) on the growth of *E. coli*. They found that transferrin saturation increased and non-transferrin-bound iron (NTBI) appeared only after transfusion with older RBCs, and that the increase in NTBI correlated with enhanced proliferation of *E. coli in vitro* (Hod et al., 2011).

In a mouse model, Prestia et al. found that transfusion of older RBCs exacerbated infection with both Gram-negative intracellular (*Salmonella typhimurium*) and extracellular (*Escherichia coli*) pathogens, and the administration of iron dextran exacerbated *E. coli* infection. They also found that co-infection with *Plasmodium yoelli* and *Salmonella typhimurium* produced overwhelming *Salmonella* sepsis. The authors suggested that iron delivery to macrophages was the mechanism that was mediating the adverse effects of iron administration (either through transfusion or supplementation) (Prestia et al., 2014). Portugal et al. also used a mouse model to study malaria "superinfections", which occur when an individual who already has blood-stage malaria parasitaemia is reinfected with liver-tropic sporozoites by a *Plasmodium*-infected mosquito. They hypothesized that superinfections would elevate the risk of hyperparasitaemia and life-threatening episodes of disease; however, they found that ongoing blood-stage infections, above a minimum threshold, actually impaired the growth of subsequently inoculated sporozoites such that they became growth-arrested and failed to develop into blood-stage parasites. They also found that the synthesis of hepcidin (an iron regulatory hormone) was stimulated by blood-stage parasites in a density-dependent manner, indicating that current parasitaemia may protect against additional malaria superinfection via the up-regulation of hepcidin and redistribution of iron away from infective competitors. The authors concluded by suggesting that altering iron availability (e.g. with iron supplementation) could increase malaria infection severity by disturbing an intrinsic balance between host iron levels, blood parasitaemia, and the hepcidin response, though further research was needed to confirm these relationships and interactions in

humans living in malaria endemic regions (Portugal et al., 2011). While the analyses from this thesis cannot confirm the findings from these recent *in vitro* and animal studies, they do support a potential relationship between host iron status and the susceptibility or proliferation of malaria and non-malaria infections among children in a malaria endemic area that varies across space and may be influenced by environmental factors or other geographical characteristics.

The second research pillar of the Iron and Malaria project pertained to the assessment and identification of *biomarkers* that can accurately and reliably measure nutritional iron deficiency in areas of endemic infection. Due to the strong up-regulating effect of inflammation on serum ferritin concentration, assessing iron status in a population where infection is prevalent can be very challenging (WHO, 2011a). I corrected ferritin for the effect of inflammation using the regression method, which was developed by the BRINDA study working group using 15 country datasets (representing over 25,000 children) (P.S. Suchdev et al., 2016). The regression method can be used to correct ferritin concentration for either CRP or AGP or both, and has been shown to provide more accurate estimates of iron status compared to other methods that rely on imposing CRP or AGP cut-offs (D. I. Thurnham et al., 2010; WHO, 2011a). Findings from my spatial analyses indicated that using the regression method with CRP only may not have been sufficient for correcting ferritin for infection in the Ghana trial setting. Since CRP reaches its peak concentration during the early phase of an inflammatory response (approximately 48 hours), while ferritin can remain elevated well into the late phase (2-8 days) (D.I. Thurnham & McCabe, 2012), it is possible that I was limited in the breadth of non-malaria infections that I could account for. Therefore, correcting ferritin for more than one biomarker of inflammation (e.g. CRP and AGP) should be recommended. Further, since the spatial model outputs suggested that malaria parasitaemia contributed to the variation in serum ferritin concentration, it may be appropriate to correct ferritin for malaria as well as inflammation. Our additional contribution to the biomarker pillar is the introduction of a new type of indicator, "geo-indicators", that could be considered as a complementary measure for assessing the iron deficiency or infection risk profile of populations in malaria endemic areas that is more feasible and less costly to measure compared to biological markers.

The contribution of this dissertation to the third pillar is, perhaps, more subtle. The goal of this research priority was to evaluate the safety and effectiveness of *interventions* to prevent and treat iron deficiency. Since the Pemba trial, research in this area has investigated the interaction between iron status and malaria transmission, the safety and efficacy of iron supplementation in pregnant women, and the impact of iron supplementation and fortification with various dosing regimens and forms among women and children in both developed and developing countries. Frosch et al. conducted a prospective cohort study of 190 children (4-59 months) living in the highland areas of Kenya to determine whether iron status improved in children living in 2 villages with a documented 12-month cessation in malaria transmission. Over the 12-month period, they found that mean Hb and ferritin increased significantly, while median CRP remained low (<1 mg/L), and iron deficiency prevalence (ferritin <12 mcg/L or <30 mcg/L if CRP ≥10 mg/L) decreased (from 35.9% to 24.9%). The authors concluded that, in an area with higher elevation and unstable malaria transmission, the prevalence of iron deficiency in children may be decreased through the interruption of malaria transmission, and that malaria elimination strategies themselves may be an effective way to improve iron status in malaria-endemic areas (Frosch et al., 2014). Malaria control was also named as a contributing factor to the success of an iron supplementation intervention in pregnant women. Etheredge et al. conducted a randomized, double-blind, placebo-controlled clinical trial to evaluate the safety and efficacy of iron supplementation among iron-replete, non-anaemic pregnant women in Dar es Salaam, Tanzania. The women (n=1500) received either 60 mg of supplemental iron or placebo and were followed up every 4 weeks as part of standard prenatal care (including malaria screening and prophylaxis and treatment as needed). They found that iron supplementation significantly improved the increase in haemoglobin and serum ferritin from baseline to delivery, and did not increase the risk of placental malaria or affect birth weight compared to placebo. The authors concluded that prenatal iron supplementation among iron-replete non-anaemic women was not associated with adverse pregnancy outcomes in the context of good malaria control (Etheredge et al., 2015).

In terms of evidence regarding the differential impacts of iron supplementation versus fortification, Brittenham et al. conducted a prospective randomized crossover study (Zurich, Switzerland) to determine the effects of iron supplementation versus fortification on the

production of circulating non-transferrin-bound iron (NTBI) in healthy women (18-39y) with replete (ferritin >25 mcg/L, n=16) and reduced iron stores (ferritin ≤25 mcg/L, n=16). The women received either a supplemental dose of 60mg iron with water, a supplemental dose of 60mg iron with a standard test meal, or a fortification dose of 6mg iron with a standard test meal. They found that serum NTBI reached peak concentrations in the supplement groups at 4 hours post-administration, though was not detectable in the fortification group. The authors concluded that the production of serum NTBI in healthy women was determined by the rate and amount of iron absorbed, which was highest when iron was administered in a supplemental dose without food and negligible when administered in a fortification dose with a meal (G. M. Brittenham et al., 2014).

As described by Hurrell, NTBI has been hypothesized to increase the intensity of malaria infection partly through the sequestration of malaria-infected red blood cells in the capillaries of the intestine, leading to increased permeability of the intestinal barrier to pathogens (R. F. Hurrell, 2011). In a high permeability state, the immune function of gut bacteria could be further impaired through frequent high doses of supplemental iron by stimulating the growth of pathogenic bacteria in the intestine (R. Hurrell, 2010). This hypothesis was tested by Dostal et al. who conducted a randomized, placebo-controlled intervention trial to investigate the impact of iron supplementation on the abundance of dominant bacterial groups in the gut, faecal short chain fatty acid (SCFA) concentration and gut inflammation among iron deficient children (6-11 years old) living in rural South Africa. The children received either oral iron tablets (50 mg ferrous sulfate) (n=22) or placebo (n=27) for 4d/week or no treatment (n=24) for 38 weeks. They found that iron supplementation improved iron status though did not significantly increase calprotectin concentration (a measure of intestinal inflammation). They also did not find a significant effect of iron supplementation on the concentrations of bacterial groups in the gut or faecal SCFA compared to placebo (Dostal et al., 2014).

These findings may be unexpected considering the hypothesized adverse effects of iron on the gut microbiome and immune function; however, the study population assessed by Dostal et al. consisted of only iron deficient children, which would have removed the potential mediating

effect of iron status. Further, it is not known whether iron fortification would have produced similar results, or whether the older age of the children was a contributing factor to the outcomes observed. Jaeggi et al. began addressing these gaps by conducting two double-blind randomized controlled trials to determine the effect of high and low dose in-home iron fortification on the gut microbiome and intestinal inflammation of Kenyan infants (6 months of age) with varying levels of iron sufficiency. The children (n=119) received maize porridge fortified with micronutrient powders that contained either 2.5 mg iron as NaFeEDTA or 12.5 mg iron as ferrous fumarate or no iron for 4 months. They found that the provision of iron-containing MNPs adversely affected the gut microbiome, leading to increased pathogen abundance and intestinal inflammation. The increase in fecal calprotectin from iron fortification was significant among those who were iron-sufficient at baseline though not among those who were iron deficient. There was also a (non-significant) higher proportion of children in the 12.5 mg iron group (27.3%) who required treatment for diarrhoea compared to those in the no-iron group (8.3%) (Jaeggi et al., 2015).

Similar to the study in Kenya, the Ghana trial (from which data were obtained for this dissertation) included a randomized intervention of iron-containing micronutrient powders (with 12.5 mg of iron as ferrous fumarate) to assess the impact of iron fortification on malaria incidence in young Ghanaian children (6-35 months of age). Zlotkin et al. found that the iron intervention did not have an adverse effect on clinical malaria incidence; however there were increased rates of hospitalization in the Iron group (Zlotkin et al., 2013). The analyses presented here added a spatial perspective to the Ghana trial outcomes, and did not show differential impacts of the intervention on infection status between groups, whether it was defined using parasitaemia or CRP. The effect of the intervention on iron status was more difficult to interpret and compare to the current literature, partly due to the confounding effect of infection (R.F. Hurrell, 2012), though also because of the lack of studies that used serum ferritin concentration as a continuous outcome variable with correction for inflammation. Despite these limitations, the maps of the predicted mean and residual variation revealed a change in the spatial patterns of both iron and infection status from baseline to endline, suggesting that geographical factors may also be important to consider when evaluating the safety and effectiveness of iron interventions in malaria endemic areas.

7.3 Limitations

A potential drawback of using spatial data, particularly satellite-derived variables, is that they often vary at a higher level than the individual (e.g. village or region) and thus can increase the risk of a change of support problem (COSP). The COSP is concerned with situations when the spatial process of interest is of one form but the data observed are of another form, resulting in a transformation of the process of interest (Gotway & Young, 2002). The support of a variable refers to the size or volume associated with its individual values, and thus averaging or aggregating a variable changes its support along with its statistical and spatial properties (Gotway & Young, 2002). Therefore, the COSP relates to the issue of how the spatial variation of one variable with a given support is associated with another with a different support. It is related to the concept of ecological fallacy, which can occur when the results based on aggregate (zonal) data are applied to individuals or specific geographical coordinates (Dark & Bram, 2007). There are several types of COSPs, and proposed solutions for each can be found throughout the statistical literature. Of particular relevance to this project is the "area-to-point" COSP, as satellite-derived data with spatial resolutions ranging from 90 to 500 meters were used to model individual-level outcomes. A potential solution to this COSP is to use Bayesian hierarchical models to accommodate the different spatial scales of the variables (Gotway & Young, 2002). This is essentially what a geostatistical model does by including a spatial random effect (P. Diggle & Ribeiro, 2007). Adjusting the spatial resolution or using different levels of aggregation or areal unit definitions are other potential ways of assessing the robustness of the associations observed to changes in scale.

The effect of the iron intervention (in the Iron group) and malaria treatment (across both groups) may have reduced the ability to gain insight into potential sources of spatial variation, specifically at endline. To my knowledge, there are very few examples in the published literature of spatial analyses investigating the geographical distribution of nutrition-related outcomes among children after a targeted intervention; however, certain parallels may be drawn

from examples outside of the nutrition literature. For example, Root et al. conducted a secondary spatial analysis of data from a randomized trial examining the efficacy of a pneumococcal conjugate vaccine among children less than 2 years of age in Bohol, Philippines (Root et al., 2014). They included 11,500 children in the per-protocol analysis and found that vaccine efficacy increased with distance to the main study hospital (approximately 4.5% per kilometer) (Root et al., 2014). Although the secondary spatial analyses of data from the Ghana trial showed a change in the associations of iron or infection status from baseline to endline, it was difficult to determine the relative contribution of treatment effects or potential confounding due to interactions with host status at baseline.

Another potential limitation of the present analyses was the incomplete correction of serum ferritin for inflammation or infection. C-reactive protein (CRP) rises in accordance with the early phase of the inflammatory response (approximately 0-2 days), while other acute phase proteins, such as alpha-1-acid glycoprotein (AGP), reach their peak concentration during the late phase (approximately 2-8 days) (D.I. Thurnham & McCabe, 2012). Since CRP was the only inflammatory biomarker available for the current analysis, and the study protocol was not designed to include diagnoses of non-malarial infections, it is possible that the prevalence of infection among participants was underestimated, which may have also lead to the overestimation of iron status at baseline and endline. Other than clinical malaria (parasitaemia with fever), the outcome measures used in the endline infection analyses were only available at baseline and endline. As such, infection status at baseline was used as an individual-level bio-indicator for general health status and a means of controlling for one's susceptibility to subsequent infections. This analytical practice is not optimal, as individual differences in immune status or antibody production in response to parasitic infections are likely (Ashton et al., 2015) and may be modulated by other infection-specific or sensitive risk factors. Although an acceptable solution to this issue does not seem to exist, the consideration of geographical factors and spatial relationships may help to fill this gap.

Lastly, an additional limitation was the use of Euclidean distance rather than network distance when estimating proximity to a health facility or the district centre. Network distance may have

been a more appropriate indicator of travel impedance or access since it can more appropriately account for travel distance by vehicle or bicycle. Calculating distance by road from all study compounds to the nearest health facility or District Centre was not possible due to incomplete or missing vector information (e.g. miss-aligned junctions, missing or disconnected road segments). We considered using the assumption that all travel distances started at the nearest road, though this was not appropriate for those villages in more remote locations with highly dispersed compounds. Nesbitt et al. (2014) encountered similar challenges in a study comparing different measures of travel impedance to estimate access to delivery care in the Brong-Ahafo region of Ghana (Nesbitt et al., 2014). After addressing each challenge and creating a detailed network map layer, the authors found that straight-line distance was as informative as network distance for determining spatial access in this setting of rural Ghana (Nesbitt et al., 2014). Therefore, I concluded that using Euclidean rather than network distance in the present analysis could be justified.

7.4 Conclusions & directions

The current analyses are the first of their kind to use model-based geostatistics with Bayesian inference and INLA to explore the spatial dynamics of iron status and malaria and non-malaria infection risk among children in rural Ghana. They also provided a glimpse into the potential utility of "geo-indicators", particularly for assessing infection risk potential, which could help guide the implementation of iron interventions in areas where additional infection control support is needed. Future research should focus on refining the spatial models (and thus improving the validity of geo-indicators) by including biomarkers of non-malaria infection that represent both the early and late stages of the acute phase response. These methods and refined models could then be applied to spatial datasets from other Regions in Ghana, or perhaps other malaria endemic countries, in order to further extend their utility. Until then, the importance of considering geographical distribution when assessing the risk of iron deficiency or infection among children in malaria endemic areas should continue to be emphasized.

References (Chapters 1-3 & 7)

Adish, A. A., Esrey, S. A., Gyorkos, T. W., & Johns, T. (1999). Risk factors for iron deficiency anaemia in preschool children in northern Ethiopia. *Public Health Nutrition, 2*(3), 243-252.

Aimone, A. M., Perumal, N., & Cole, D. C. (2013). A systematic review of the application and utility of geographical information systems for exploring disease-disease relationships in paediatric global health research: the case of anaemia and malaria. *International Journal of Health Geographics, 12*, 1-13.

Alegana, V. A., Atkinson, P. M., Wright, J. A., Kamwi, R., Uusiku, P., Katokele, S., . . . Noor, A. M. (2013). Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models. *Spat Spatiotemporal Epidemiol, 7*, 25-36. doi:10.1016/j.sste.2013.09.001

Andrews, N. C., & Schmidt, P. J. (2007). Iron homeostasis. *Annu Rev Physiol, 69*, 69-85. doi:10.1146/annurev.physiol.69.031905.164337

Ashton, R. A., Kefyalew, T., Rand, A., Sime, H., Assefa, A., Mekasha, A., . . . Brooker, S. J. (2015). Geostatistical modeling of malaria endemicity using serological indicators of exposure collected through school surveys. *Am J Trop Med Hyg, 93*(1), 168-177. doi:10.4269/ajtmh.14-0620

Balarajan, Y., Ramakrishnan, U., Ozaltin, E., Shankar, A. H., & Subramanian, S. V. (2011). Anaemia in low-income and middle-income countries. *Lancet, 378*(9809), 2123-2135. doi:10.1016/S0140-6736(10)62304-5

Beard, J. L. (2001). Iron biology in immune function, muscle metabolism and neuronal functioning. *J Nutr, 131*(2S-2), 568S-579S; discussion 580S.

Bellhouse, D. R. (2004). The Reverend Thomas Bayes, FRS: A biography to celebrate the tercentenary of his birth. *Statistical Science, 19*(1), 3-43.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., . . . Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature, 526*(7572), 207-211. doi:10.1038/nature15535

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., . . . Rivera, J. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet, 371*(9608), 243-260.

Blangiardo, M., Cameletti, M., Baio, G., & Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spat Spatiotemporal Epidemiol, 7*, 39-55.

Brabin, L., Brabin, B. J., & Gies, S. (2013). Influence of iron status on risk of maternal or neonatal infection and on neonatal mortality with an emphasis on developing countries. *Nutr Rev, 71*(8), 528-540. doi:10.1111/nure.12049

Brink, E. W., Aly, H. E., Dakroury, A. M., Said, A. K., Ghoneme, F. M., Hussein, M. A., . . . Nichaman, M. Z. (1983). The Egyptian National Nutrition Survey, 1978. *Bull World Health Organ, 61*(5), 853-860.

Brittenham, G. M. (2012). *Safety of iron fortification and supplementation in malaria-endemic areas*. Paper presented at the Nestle Nutrition Institute Workshop Series, Basel.

Brittenham, G. M., Andersson, M., Egli, I., Foman, J. T., Zeder, C., Westerman, M. E., & Hurrell, R. F. (2014). Circulating non-transferrin-bound iron after oral administration of supplemental and fortification doses of iron to healthy women: a randomized study. *Am J Clin Nutr, 100*(3), 813-820. doi:10.3945/ajcn.113.081505

Brooks, S., Gelman, A., Jones, G., & Meng, X. (2011). *Handbook of Markov chain Monte Carlo*: CRC Press, Taylor & Francis Group.

Brown, P. E. (2015). Model-based geostatistics the easy way. *Journal of Statistical Software, 63*(12), 1-24.

Buechner, J. S., Constantine, H., & Gjelsvik, A. (2004). John Snow and the Broad Street pump: 150 years of epidemiology. *Med Health R I, 87*(10), 314-315.

Clark, M., Fisher, N. C., Kasthuri, R., & Cerami Hand, C. (2013). Parasite maturation and host serum iron influence the labile iron pool of erythrocyte stage Plasmodium falciparum. *Br J Haematol, 161*(2), 262-269. doi:10.1111/bjh.12234

Clark, M. A., Goheen, M. M., & Cerami, C. (2014). Influence of host iron status on Plasmodium falciparum infection. *Front Pharmacol, 5*, 84. doi:10.3389/fphar.2014.00084

Clark, M. A., Goheen, M. M., Fulford, A., Prentice, A. M., Elnagheeb, M. A., Patel, J., . . . Cerami, C. (2014). Host iron status and iron supplementation mediate susceptibility to erythrocytic stage Plasmodium falciparum. *Nat Commun, 5*, 4446. doi:10.1038/ncomms5446

Collaborators, G. B. o. D. S. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet, 386*(9995), 743-800. doi:10.1016/S0140-6736(15)60692-4

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography, 31*(5), 471-479.

Davis, J. C. (1972). *Statistics and Data Analysis in Geology* (2nd ed.). New York: Wiley.

De-Regil, L. M., Suchdev, P. S., Vist, G. E., Walleser, S., & Pena-Rosas, J. P. (2011). Home fortification of foods with multiple micronutrient powders for health and nutrition in children under two years of age. *Cochrane database of systematic reviews (Online), 9*, CD008959.

Diggle, P., Moyeed, R., Rowlingson, B., & Thomson, M. (2002). Childhood malaria in the Gambia: A case-study in model-based geostatistics. *Journal of the Royal Statistical Society, 51*(4), 493-506.

Diggle, P., & Ribeiro, P. J. (2007). *Model-based geostatistics*. New York, NY: Springer.

Diggle, P. J., Moyeed, R. A., & Tawn, J. A. (1998). Model-based geostatistics. *Applied Statistics, 47*, 299-350.

Diggle, P. J., & Ribeiro, P. J. (2006). *Model-based geostatistics*. New York: Springer-Verlag.

Donovan, A., Lima, C. A., Pinkus, J. L., Pinkus, G. S., Zon, L. I., Robine, S., & Andrews, N. C. (2005). The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metab, 1*(3), 191-200. doi:10.1016/j.cmet.2005.01.003

Dostal, A., Baumgartner, J., Riesen, N., Chassard, C., Smuts, C. M., Zimmermann, M. B., & Lacroix, C. (2014). Effects of iron supplementation on dominant bacterial groups in the gut, faecal SCFA and gut inflammation: a randomised, placebo-controlled intervention trial in South African children. *Br J Nutr, 112*(4), 547-556. doi:10.1017/S0007114514001160

Dunson, D. B. (2001). Commentary: practical advantages of Bayesian analysis of epidemiologic data. *Am J Epidemiol, 153*(12), 1222-1226.

ESRI. (2011). Mapping Software and Data. Retrieved from www.esri.com

Etheredge, A. J., Premji, Z., Gunaratna, N. S., Abioye, A. I., Aboud, S., Duggan, C., . . . Fawzi, W. W. (2015). Iron Supplementation in Iron-Replete and Nonanemic Pregnant Women in Tanzania: A Randomized Clinical Trial. *JAMA Pediatr, 169*(10), 947-955. doi:10.1001/jamapediatrics.2015.1480

FAO/WHO. (2005). *Joint FAO/WHO Expert Consultation on Human Vitamin, Mineral Requirements. Iron.* Retrieved from Geneva:

Fong, Y., Rue, H., & Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics, 11*(3), 397-412. doi:10.1093/biostatistics/kxp053

Frosch, A. E., Ondigo, B. N., Ayodo, G. A., Vulule, J. M., John, C. C., & Cusick, S. E. (2014). Decline in childhood iron deficiency after interruption of malaria transmission in highland Kenya. *Am J Clin Nutr, 100*(3), 968-973. doi:10.3945/ajcn.114.087114

Ganz, T. (2005). Hepcidin--a regulator of intestinal iron absorption and iron recycling by macrophages. *Best Pract Res Clin Haematol, 18*(2), 171-182. doi:10.1016/j.beha.2004.08.020

Ganz, T., & Nemeth, E. (2009). Iron sequestration and anemia of inflammation. *Semin Hematol, 46*(4), 387-393. doi:10.1053/j.seminhematol.2009.06.001

Gelfand, A., Diggle, P., M., F., & Guttorp, P. (2010). *Handbook of spatial statistics*: Chapman & Hall.

Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., & Smith, T. (2004). Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *Am J Epidemiol, 159*(1), 64-72.

Ghana. (2016). Brong-Ahafo Region. Retrieved from http://ghana.gov.gh/index.php/about-ghana/regions/brong-ahafo

Giardina, F., Gosoniu, L., Konate, L., Diouf, M. B., Perry, R., Gaye, O., . . . Vounatsou, P. (2012). Estimating the burden of malaria in Senegal: Bayesian zero-inflated binomial geostatistical modeling of the MIS 2008 data. *PLoS One, 7*(3), e32625. doi:PONE-D-11-14663 [pii] 10.1371/journal.pone.0032625

Gosoniu, L., Veta, A. M., & Vounatsou, P. (2010). Bayesian geostatistical modeling of Malaria Indicator Survey data in Angola. P*LoS One, 5(*3), e9322. doi:10.1371/journal.pone.0009322

Gotway, A., & Young, L. J. (2002). Combining incompatible spatial data. J*ournal of the American Statistical Association, 97(*458), 632-648.

Greenberg, G. R., & Ashenbrucker, H. (1947). The anemia of infection; fate of injected radioactive iron in the presence of inflammation. J *Clin Invest, 26(*1), 121-125.

Greenland, S. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. I*nt J Epidemiol, 35(*3), 765-775. doi:10.1093/ije/dyi312

GSS. (2012). 2*010 Population and Housing Census: Final Results.* Retrieved from Ghana:

Halliwell, B., & Gutteridge, J. M. C. (1989). F*ree radicals in biology and medicine.* Oxford: Clarendon Press.

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., . . . Snow, R. W. (2009). A world malaria map: Plasmodium falciparum endemicity in 2007. P*LoS Med, 6(*3), e1000048. doi:10.1371/journal.pmed.1000048

Held, L., Schrodle, B., & Rue, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In T. Kneib & G. Tutz (Eds.), S*tatistical Modelling and Regression Structures* (pp. 91-110). Heidelberg: Springer-Verlag.

Hengl, T. (2014). Worldgrids — a public repository and a WPS for global environmental layers. Retrieved from http://worldgrids.org

HF-TAG. (2011). H*ome-Fortification with Micronutrient Powders.* Retrieved from

Hod, E. A., Brittenham, G. M., Billote, G. B., Francis, R. O., Ginzburg, Y. Z., Hendrickson, J. E., . . . Spitalnik, S. L. (2011). Transfusion of human volunteers with older, stored red blood cells produces extravascular hemolysis and circulating non-transferrin-bound iron. B*lood, 118(*25), 6675-6682. doi:10.1182/blood-2011-08-371849

Hurrell, R. (2010). Iron and malaria: absorption, efficacy and safety. I*nt J Vitam Nutr Res, 80(*4-5), 279-292. doi:10.1024/0300-9831/a000035

Hurrell, R. F. (2011). Safety and efficacy of iron supplements in malaria-endemic areas. A*nn Nutr Metab, 59(*1), 64-66. doi:10.1159/000332140

Hurrell, R. F. (2012). I*nfluence of inflammatory disorders and infection on iron absorption and efficacy of iron-fortified foods.* Paper presented at the Nestle Nutrition Institute Workshop Series, Basel.

Hurt, L., ten Asbroek, A., Amenga-Etego, S., Zandoh, C., Danso, S., Edmond, K., . . . Kirkwood, B. R. (2013). Effect of vitamin A supplementation on cause-specific

mortality in women of reproductive age in Ghana: a secondary analysis from the ObaapaVitA trial. B*ull World Health Organ, 91(*1), 19-27. doi:10.2471/BLT.11.100412

IOM. (1997-2001). Dietary Reference Intakes Tables and Application. Retrieved from http://www.iom.edu/Activities/Nutrition/SummaryDRIs/DRI-Tables.aspx

Jackman, S. (2009). B*ayesian analysis for the social sciences:* Wiley-Blackwell.

Jaeggi, T., Kortman, G. A., Moretti, D., Chassard, C., Holding, P., Dostal, A., . . . Zimmermann, M. B. (2015). Iron fortification adversely affects the gut microbiome, increases pathogen abundance and induces intestinal inflammation in Kenyan infants. G*ut, 64(*5), 731-742. doi:10.1136/gutjnl-2014-307720

Jefferds, M. E., Irizarry, L., Timmer, A., & Tripp, K. (2013). UNICEF-CDC global assessment of home fortification interventions 2011: current status, new directions, and implications for policy and programmatic guidance. F*ood Nutr Bull, 34(*4), 434-443.

Jonker, F. A., Calis, J. C., van Hensbroek, M. B., Phiri, K., Geskus, R. B., Brabin, B. J., & Leenstra, T. (2012). Iron status predicts malaria risk in Malawian preschool children. P*LoS One, 7(*8), e42670. doi:10.1371/journal.pone.0042670

Lindgren, F., Rue, H., & Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J*ournal of the Royal Statistical Society (Series B, Statistical Methodology), 73, Part 4,* 423-498.

Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T., & Murray, C. J. (2006). Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. L*ancet, 367(*9524), 1747-1757. doi:10.1016/S0140-6736(06)68770-9

Mainardi, S. (2012). Modelling spatial heterogeneity and anisotropy: child anaemia, sanitation and basic infrastructure in sub-Saharan Afrca. I*nternational Journal of Geographical Information Science, 26(*3), 387-411.

Masters, S. H., Burstein, R., Amofah, G., Abaogye, P., Kumar, S., & Hanlon, M. (2013). Travel time to maternity care and its effect on utilization in rural Ghana: a multilevel analysis. S*oc Sci Med, 93,* 147-154. doi:10.1016/j.socscimed.2013.06.012

Mei, Z., Cogswell, M. E., Parvanta, I., Lynch, S., Beard, J. L., Stoltzfus, R. J., & Grummer-Strawn, L. M. (2005). Hemoglobin and ferritin are currently the most efficient indicators of population response to iron interventions: an analysis of nine randomized controlled trials. J *Nutr, 135(*8), 1974-1980.

Menendez, C., Quinto, L. L., Kahigwa, E., Alvarez, L., Fernandez, R., Gimenez, N., . . . Alonso, P. L. (2001). Effect of malaria on soluble transferrin receptor levels in Tanzanian infants. A*m J Trop Med Hyg, 65(*2), 138-142.

Murray, M. J., Murray, A. B., Murray, M. B., & Murray, C. J. (1978). The adverse effect of iron repletion on the course of certain infections. B*r Med J, 2(*6145), 1113-1115.

Nemeth, E., Tuttle, M. S., Powelson, J., Vaughn, M. B., Donovan, A., Ward, D. M., . . . Kaplan, J. (2004). Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. S*cience, 306(*5704), 2090-2093. doi:10.1126/science.1104742

Nesbitt, R. C., Gabrysch, S., Laub, A., Soremekun, S., Manu, A., Kirkwood, B. R., . . . Grundy, C. (2014). Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. I*nt J Health Geogr, 13,* 25. doi:10.1186/1476-072X-13-25

Nykiforuk, C., & Flaman, L. (2008). E*xploring the utilization of geographic information systems in health promotion and public health.* Retrieved from Edmonton, AB, Canada:

Ojukwu, J. U., Okebe, J. U., Yahav, D., & Paul, M. (2009). Oral iron supplementation for preventing or treating anaemia among children in malaria-endemic areas. C*ochrane Database Syst Rev(*3), CD006589. doi:10.1002/14651858.CD006589.pub2

Okebe, J. U., Yahav, D., Shbita, R., & Paul, M. (2011). Oral iron supplements for children in malaria-endemic areas. C*ochrane Database Syst Rev(*10), CD006589. doi:10.1002/14651858.CD006589.pub3

Oppenheimer, S. (2001). Iron and its relation to immunity and infectious disease. J *Nutr, 131,* 616S-635S.

Paul, M., Riebler, A., Bachmann, L. M., Rue, H., & Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. S*tat Med, 29(*12), 1325-1339. doi:10.1002/sim.3858

Portugal, S., Carret, C., Recker, M., Armitage, A. E., Gonçalves, L. A., Epiphanio, S., . . . Mota, M. M. (2011). Host-mediated regulation of superinfection in malaria. N*at Med, 17(*6), 732-737. doi:10.1038/nm.2368

Prentice, A. M., Verhoef, H., & Cerami, C. (2013). Iron fortification and malaria risk in children. J*AMA, 310(*9), 914-915. doi:10.1001/jama.2013.6771

Prestia, K., Bandyopadhyay, S., Slate, A., Francis, R. O., Francis, K. P., Spitalnik, S. L., . . . Hod, E. A. (2014). Transfusion of stored blood impairs host defenses against Gram-negative pathogens in mice. T*ransfusion, 54(*11), 2842-2851. doi:10.1111/trf.12712

R. (2015). R: A Language and Environment for Statistical Computing (Version 3.2.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Raiten, D., Namaste, S., & Brabin, B. (2009). C*onsiderations for the safe and effective use of iron interventions in areas of malaria burden: Full technical report.* Retrieved from USA:

Raiten, D. J., & Combs, G. F. J. (2015). Directions in Nutritional Assessment: Biomarkers and bio-indicators: providing clarity in the face of complexity. S*ight and Life, 29(*1), 39-44.

Raiten, D. J., Namasté, S., & Brabin, B. (2011). Considerations for the safe and effective use of iron interventions in areas of malaria burden - executive summary. I*nt J Vitam Nutr Res, 81(*1), 57-71. doi:10.1024/0300-9831/a000051

Raiten, D. J., Sakr Ashour, F. A., Ross, A. C., Meydani, S. N., Dawson, H. D., Stephensen, C. B., . . . Group, I. C. (2015). Inflammation and Nutritional Science for Programs/Policies

and Interpretation of Research Evidence (INSPIRE). J *Nutr, 145(*5), 1039S-1108S. doi:10.3945/jn.114.194571

Raso, G., Schur, N., Utzinger, J., Koudou, B. G., Tchicaya, E. S., Rohner, F., . . . Vounatsou, P. (2012). Mapping malaria risk among children in Côte d'Ivoire using Bayesian geo-statistical models. M*alar J, 11,* 160. doi:10.1186/1475-2875-11-160

Riebler, A., Held, L., & Rue, H. (2012). Estimation and extrapolation of time trends in registry data - borrowing strength from related populations. A*nn Appl Stat, 6(*1), 304-333.

Root, E. D., Lucero, M., Nohynek, H., Anthamatten, P., Thomas, D. S., Tallo, V., . . . Simões, E. A. (2014). Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children. P*roc Natl Acad Sci U S A, 111(*9), 3520-3525. doi:10.1073/pnas.1313748111

Rue, H., & Held, L. (2005). G*aussian Markov random fields. Theory and applications:* Chapman & Hall.

Rue, H., & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. J *Stat Plann Infer, 137,* 3177-3192.

Rue, H., & Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J*ournal of the Royal Statistical Society: Series B, 71(*2), 319-392.

Samadoulougou, S., Maheu-Giroux, M., Kirakoya-Samadoulougou, F., De Keukeleire, M., Castro, M. C., & Robert, A. (2014). Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. P*arasit Vectors, 7,* 350. doi:10.1186/1756-3305-7-350

Sazawal, Z., Black, RE., Ramsan, M., Chwaya, HM., Stoltzfus, RJ., Dutta, A ., & et al. (2006). Effects of routine prophylactic supplementation with iron and folic acid on admission to hospital and mortality in preschool children in a high malaria transmission setting: community-based, randomised, placebo-controlled trial. L*ancet, 367,* 133-143.

Schrodle, B., & Held, L. (2011). Spatio-temporal disease mapping using INLA. E*nvionmetrics, 22(*6), 725-734.

Schur, N., Vounatsou, P., & Utzinger, J. (2012). Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. P*LoS Negl Trop Dis, 6(*9), e1773. doi:10.1371/journal.pntd.0001773

Shirayama, Y., Phompida, S., & Shibuya, K. (2009). Geographic information system (GIS) maps and malaria control monitoring: intervention coverage and health outcome in distal villages of Khammouane province, Laos. M*alar J, 8,* 217. doi:10.1186/1475-2875-8-217

Snow, R. W., Bastos de Azevedo, I., Lowe, B. S., Kabiru, E. W., Nevill, C. G., Mwankusye, S., . . . Teuscher, T. (1994). Severe childhood malaria in two areas of markedly different falciparum transmission in east Africa. A*cta Trop, 57(*4), 289-300.

Soares Magalhães, R. J., & Clements, A. C. (2011). Spatial heterogeneity of haemoglobin concentration in preschool-age children in sub-Saharan Africa. B*ull World Health Organ, 89(*6), 459-468. doi:10.2471/BLT.10.083568

Soares Magalhães, R. J., Langa, A., Pedro, J. M., Sousa-Figueiredo, J. C., Clements, A. C., & Vaz Nery, S. (2013). Role of malnutrition and parasite infections in the spatial variation in children's anaemia risk in northern Angola. G*eospat Health, 7(*2), 341-354. doi:10.4081/gh.2013.91

Song, P., Zhu, Y., Mao, X., Li, Q., & An, L. (2013). Assessing spatial accessibility to maternity units in Shenzhen, China. P*LoS One, 8(*7), e70227. doi:10.1371/journal.pone.0070227

Stevens, G. A., Finucane, M. M., De-Regil, L. M., Paciorek, C. J., Flaxman, S. R., Branca, F., . . . (Anaemia), N. I. M. S. G. (2013). Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data. L*ancet Glob Health, 1(*1), e16-25. doi:10.1016/S2214-109X(13)70001-9

Stoltfus, R. J., Mullany, L., & Black, R. E. (2004). Iron deficiency anaemia. In M. L. Ezzati, A.D.Rodgers, A.Murray, C.J.L. (Ed.), C*omparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors* (Vol. 1). Geneva, Switzerland: World Health Organization.

Stoltzfus, R. J. (2003). Iron deficiency: global prevalence and consequences. F*ood Nutr Bull, 24(*4 Suppl), S99-103.

Suchdev, P. S., Namaste, S., Aaron, G., Raiten, D. J., Brown, K. H., & Flores-Ayala, R. (2016). Overview of the Biomarkers Reflecting Inflammtion and Determinants of Anemia (BRINDA) Project. A*dvances in Nutrition, 7(*2), 349-356.

Suchdev, P. S., Ruth, L. J., Woodruff, B. A., Mbakaya, C., Mandava, U., Flores-Ayala, R., . . . Quick, R. (2012). Selling Sprinkles micronutrient powder reduces anemia, iron deficiency, and vitamin A deficiency in young children in Western Kenya: a cluster-randomized controlled trial. A*m J Clin Nutr, 95(*5), 1223-1230. doi:10.3945/ajcn.111.030072

Thomas, W. J., Koenig, H. M., Lightsey, A. L., & Green, R. (1977). Free erythrocyte porphyrin: hemoglobin ratios, serum ferritin, and transferrin saturation levels during treatment of infants with iron-deficiency anemia. B*lood, 49(*3), 455-462.

Thomson, M. C., Connor, S. J., D'Alessandro, U., Rowlingson, B., Diggle, P., Cresswell, M., & Greenwood, B. (1999). Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. A*m J Trop Med Hyg, 61(*1), 2-8.

Thurnham, D. I., Mburu, A. S., Mwaniki, D. L., & De Wagt, A. (2005). Micronutrients in childhood and the influence of subclinical inflammation. P*roc Nutr Soc, 64(*4), 502-509.

Thurnham, D. I., & McCabe, G. P. (2012). Influence of infection and inflammation on biomarkers of nutritional status with an emphasis on vitamin A and iron. In W. H. Organization (Ed.), R*eport: Priorities in the assessment of vitamin A and iron status in populations, Panama City, Panama, 15-17 September 2010* (pp. 63-80). Geneva: World Health Organization.

Thurnham, D. I., McCabe, L. D., Haldar, S., Wieringa, F. T., Northrop-Clewes, C. A., & McCabe, G. P. (2010). Adjusting plasma ferritin concentrations to remove the effects of subclinical inflammation in the assessment of iron deficiency: a meta-analysis. A*m J Clin Nutr, 92(*3), 546-555. doi:10.3945/ajcn.2010.29284

Tierney, L., & Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. J *Am Stat Assoc, 393(*81), 82-86.

USGS. (2012). United States Geological Survey. Retrieved from https://lta.cr.usgs.gov

USGS. (2014). Land Processes Distributed Active Archive Center. Retrieved from https://lpdaac.usgs.gov

van Santen, S., de Mast, Q., Swinkels, D. W., & van der Ven, A. J. (2011). Hepcidin in malaria superinfection: can findings be translated to humans? N*at Med, 17(*11), 1341; author reply 1341-1342. doi:10.1038/nm.2488

Wander, K., Shell-Duncan, B., & McDade, T. W. (2009). Evaluation of iron deficiency as a nutritional adaptation to infectious disease: an evolutionary medicine perspective. A*m J Hum Biol, 21(*2), 172-179. doi:10.1002/ajhb.20839

Weier, J., & Herring, D. (2000). Measuring Vegetation (NDVI & EVI). Retrieved from earthobservatory.nasa.gov/Features/MeasureingVegetation/measuring_vegetation_2.php

Weinberg, E. D. (1999). The role of iron in protozoan and fungal infectious diseases. J *Eukaryot Microbiol, 46(*3), 231-238.

WHO. (2006). Conclusions and recommendations of the WHO Consultation on prevention and control of iron deficiency in infants and young children in malaria endemic areas. W*orld Health Organization Secretariat on behalf of the participants to the Consultation, Lyon, France, 12–14 June(*http://www.who.int/nutrition/publications/micronutrients/FNBvol28N4supdec07.pdf), Accessed October 6, 2012.

WHO. (2009). W*orld Malaria Report 2009.* Retrieved from Geneva, Switzerland:

WHO. (2011a). Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations V*itamin and Mineral Nutrition Information System* (Vol. 11.2, pp. 1-5). Geneva: World Health Organization.

WHO. (2011b). U*se of multiple micronutrient powders for home fortification of foods consumed by infants and children 6-23 months of age.* Retrieved from Geneva, Switzerland:

WHO. (2014). W*orld Malaria Report 2014.* Retrieved from Geneva, Switzerland:

WHO. (2015). T*he global prevalence of anaemia in 2011*. Retrieved from Geneva:

WHO/CDC. (2007). Assessing the Iron Status of Populations: report of a joint WHO/CDC technical consulation on the assessment of iron status at the population level. Geneva: World Health Organization.

WHO/NHD. (2001). I*ron deficiency anaemia: assessment, prevention and control: a guide for programme managers.* Retrieved from Geneva, Switzerland:

WHO/UNICEF. World Malaria Report 2005. from http://www.rbm.who.int/wmr2005/

WHO/UNICEF. (2006). I*ron supplementation of young children in regions where malaria transmission is intense and infectious disease highly prevalent: Joint statement.* Retrieved from Geneva, Switzerland:

WHO/UNICEF. (2009). W*HO child growth standards and the identification of severe acute malnutrition in infants and children: A Joint Statement by the World Health Organization and the United Nations Children's Fund.* Retrieved from Geneva, Switzerland:

WHO/UNICEF/UNU. (2001). I*ron deficiency anaemia: assessment, prevention and control: a guide for programme managers.* Retrieved from Geneva:

Worwood, M. (2007). Indicators of the iron status of populations: ferritin. In WHO/CDC (Ed.), A*ssessing the iron status of populations: report of a joint World Health Organization/Centers for Disease Control and Prevention technical consultation on the assessment of iron status at the population level* (2nd ed.). Geneva: World Health Organization.

Zlotkin, S., Arthur, P., Antwi, K. Y., & Yeung, G. (2001). Treatment of anemia with microencapsulated ferrous fumarate plus ascorbic acid supplied as sprinkles to complementary (weaning) foods. A*m J Clin Nutr, 74(*6), 791-795.

Zlotkin, S., Arthur, P., Schauer, C., Antwi, K. Y., Yeung, G., & Piekarz, A. (2003). Home-fortification with iron and zinc Sprinkles or iron Sprinkles alone successfully treats anemia in infants and young children. J*ournal of Nutrition, 133(*4), 1075-1080.

Zlotkin, S., Newton, S., Aimone, A. M., Azindow, I., Amenga-Etego, S., Tchum, K., . . . Owusu-Agyei, S. (2013). Effect of iron fortification on malaria incidence in infants and young children in Ghana: a randomized trial. J*AMA, 310(*9), 938-947. doi:10.1001/jama.2013.277129

Appendix A

Ghana Trial Protocol – Selected Information

A.1 Sample Size

The sample size was based on the estimated incidence of clinical malaria in Ghana, defined (according to WHO criteria) as parasitaemia (malaria parasites detected on a blood smear) of any density plus history of fever (within 48 hours) or axillary temperature >37.5^{0}C. To determine treatment success and distinguish unique malaria episodes, treatments were supervised and follow-up blood smears were collected from on days 3, 7 and 14 following treatment initiation. It was hypothesize that the incidence of malaria would be significantly higher in children receiving iron versus a placebo of micronutrients without iron.

Based on the 2006 estimate of "cases and deaths from fevers suspected of being malaria" for children under 5 years of age in all of Ghana[1], the baseline rate of 3.44 episodes/child/year was assumed. It was estimated that a total of 351 person-years would be required to detect a 15% increase in malaria incidence rates with 80% power and at a 5% type I error. Considering that all children enrolled in the trial began the 5-month intervention period at a similar risk level, and accounting for 15% loss to follow-up, as well as bed net use (25% reduction), and the need to test the hypothesis twice (due to the interim analysis by the Data Safety and Monitoring committee), the sample size was calculated as 1940 children (970 per group).

A.2 Randomization Process and Blinding

Eligible children were enrolled and randomized (at the compound level), using a computer-generated model, to either the iron or No-iron group. Sachets containing powdered micronutrients with and without iron looked and tasted identical with the exception that each package was marked with an 'A' or 'B' denoting packages with or without iron. The study team

[1] WHO. World Malaria Report 2008. Geneva, Switzerland; 2008.

and caretakers of children were blinded to the ‘A’ or ‘B’ designation. The key was revealed after the database was closed and statistical analyses were completed.

A.3 Supply and content of the micronutrient powder (MNP)

The powdered mineral and vitamin fortificant product used (“Sprinkles”) had a formulation that consisted of a specific combination of minerals and vitamins, including iron in the form of microencapsulated ferrous fumarate. The product was procured from a production facility in India (Hexagon Nutrition, Mumbai, India), which has supplied a reliable high quality product for past research projects using MNPs. The production facility in India is a UNICEF-approved MNP manufacturing facility. The dose amounts of each micronutrient, including iron were: 12.5 mg iron (ferrous fumarate), 400 µg vitamin A (retinol equivalents), 30 mg vitamin C (ascorbic acid), and 5 mg zinc. iron is shown in Table 2 below. The dose of individual nutrients was based on WHO or IOM-DRI guidelines for children under 2 years of age.

A.4 Laboratory Analyses

All laboratory analyses were conducted at the Kintampo Health Research Centre by qualified laboratory technicians.

A.4.1 Complete blood count (CBC)

Blood samples were analyzed for full blood counts (CBC) using a haematology auto-analyzer (Horiba ABX Micros 60-OT). This machine requires only a few micro-litres of blood to perform analyses for haemoglobin, and various red blood cell (RBC) indices. Capillary blood was collected (via finger prick) and processed within 12 hours after collection (or 24 hours if refrigerated). The analyzer was calibrated approximately every quarter (3 months) and quality control checks were performed before running subject blood samples. The reagents used for test samples included ABX Minidil LMG, ABX Miniclean LMG, ABX Minilyse, and control blood

(Low, Normal, High) was used for quality control checks. The reference range used for haemoglobin was 11.0 – 16.5 g/dl.

A.4.2 Serum ferritin

Serum ferritin concentration was measured using the Spectro Ferritin (S-22) enzyme immunoassay. The S-22 enzyme immunoassay is a two stage reaction. In Stage 2, the addition of potassium ferricyanide causes a colour to develop, the optical density (490-510nm) of which is directly proportional to the ferritin concentration in the sample. After blood was collected in EDTA tubes, the plasma was separated from the blood cells as soon as possible and stored at -20 degrees Celsius until further analysis. Samples were thawed, mixed, brought to room temperature (18-25 degrees C), and centrifuged prior to testing. The ferritin assay was calibrated using a prediluted calibration Solutions. This solution consisted of human spleen ferritin calibrated to concentrations of 6, 20, 60, 200, 600, and 2000 ng/mL, against WHO reference material, in phosphate buffered saline with rabbit serum and sodium azide as a preservative. Quality control was performed through automatic background subtraction. If was not available, then the test sample plate was read a second time at 600-630nm and these values were manually subtracted from the initial 500nm readings. 96-well microplates were prepared using all necessary standards and controls (Prediluted Calibrator Solution and Conjugated Antihuman Ferritin). Before reading the prepared plate, the colour of the Substrate Solution was developed by adding 0.24% Potassium Ferricyanide. The absorbance of all samples was read at 490 nm using the MRX microplate reader.

A.4.3 Plasma C-reactive protein (CRP)

Plasma CRP was assayed using the QuikRead 101 CRP instrument, which is an immunoturbidimetric test based on microparticles that react with the CRP present in plasma samples. The resultant change in the turbidity of the solution is measured by the QuikRead instrument. After blood was collected in EDTA tubes, the plasma was separated from the blood

cells as soon as possible to and stored at -20 degrees Celsius until further analysis. Samples were thawed, mixed, brought to room temperature (18-25 degrees C), and centrifuged prior to testing. Quality control procedures were performed daily using the QuikRead CRP Control sample. Each assay tests required 120 µL of plasma and a CRP reagent. The lower detectable limit of the assay was 8 mg/L, meaning that values below this level were considered to be 0 mg/L. In order to obtain values between 0 and 8 mg/L, the sample concentration was increased by a factor, which was then used as a divisor for the measured value. For example, if the sample concentration was increased by a factor of 10 and the reading of the increased sample was 8 mg/L then the actual value was reported as 0.8 mg/L.

A.4.4 Malaria parasitaemia

Screening for malaria was performed in the field using a rapid diagnostic test (Paracheck Pf®) to help decide on treatment and blood smears for microscopy to provide parasite counts. For microscopy, two slides were prepared with thick and thin smears on each, using a standard template and fixed volumes (2 µL for thin smears and 6µL for thick smears). The thin smears were dried and fixed with methanol. Both slides were stained with a 3% Giemsa solution for 45 minutes, then one slide was read by 2 independent microscopists. A third reader came in if the two results were discordant. The second slide was kept as backup.

White and red blood cell counts (obtained from the haematology auto-analyzer) were used in the parasite density calculations. If 100 fields on the thick blood smear were free of parasites the slide was declared negative. *High* parasitaemia was defined as 100 parasites or more on the first field of the thick smear. In these cases, the RBC and parasitized red blood cells were then counted on the thin film. If at least 20 parasitized RBCs were counted in one field the high parasitaemia was confirmed. *Medium* parasitaemia was defined as less than 100 parasites per 200 white blood cells (WBC) within a field of the thick smear. *Low* parasitaemia was defined as less than 10 parasites per 200 WBC within a field of the thick smear, in which case the count was extended up to 500 WBC.

A.5 Clinical procedures

At *baseline* each child's weight and length measurements were recorded using procedures outlined in section A.5.1, as well as any relevant demographic (age, gender) and historical health information (birth weight and previous hospitalizations). Caregivers will also be asked to provide information regarding their usual feeding practices and health-seeking behaviours, and household information related to assets, occupation, and education (see Nutrition and Demographic data collection form in Appendix B).

A.5.1 Anthropometric measures

A.5.1.1 Body length

Length measurements were taken twice (by the same field worker) using a wooden length board with a fixed head piece and move-able foot piece. Before measuring a child's length, his/her clothes, shoes and socks were removed, as well as braids and hair ornaments. The length board was placed on a flat and stable surface. The mother was asked to lay the child on his/her back with the head against the fixed headboard, compressing the hair and ensuring that an imaginary line from the ear canal to the lower border of the eye socket was perpendicular to the board and the child's eyes were looking straight up. The field worker checked the infant's alignment, then performed the measurement by holding down the child's legs and pushing the footboard against the child's feet. Length measurements were recorded to the last completed 0.1 cm.

A.5.1.2 Body weight

Weight measurements were taken twice (by the same field worker) using a digital scale. Before measuring a child's weight, his/her clothes, shoes and socks were removed. If the child could not stand still on the scale by him/herself then the mother was asked to step on the scale with the child and her weight was subtracted through an initial taring step. Weight measurements were recorded to the nearest 100 grams.

Appendix B

Demographic and Nutrition Data Collection Form

KINTAMPO HEALTH RESEARCH CENTRE COUNTRY CODE: 1 Sprinkles Iron and Malaria Study DEMOGRAPHIC & NUTRITION FORM	DEM&NUTR Form No.:

1.0 BASIC DATA

1.1 Group					**1. A**	**2. B**	
1.2 Compound Number							
1.3 Child's name							
1.4 Child's ID Number	**S**	**P**					
1.5 Child Initials							
1.6 Respondent's name							
1.7 Respondent	1. Mother	2. Caregiver	3. Other (Specify):				
1.8 Week Number							
1.9 Date of Visit	dd/mm/yy						
1.10 Staff Code:							

2.0 CHARACTERISTICS OF RESPONDENT

2.1 Relationship of respondent to child	1. Mother	2. Other (specify):		
2.2 Age of the respondent	1.15-17 yrs	2. 18-29 yrs	3. 30-69 yrs	4. ≥70 yrs
2.3 Highest education level reached	1. None	2. Primary school		

3. Middle/continuation school, JSS	4. Technical, commercial, SSS, secondary school
5. Post-middle college: Teacher training, secretarial	6. Post secondary – nursing, teacher, polytechnic, etc.
7. University	8.Other (specify):

3.0 HOUSEHOLD CHARACTERISTICS

3.1 How many people live in your household now? | | |

3.2 How many children in your household are under 5 yrs of age? | | |

3.3 Who is the household head?

1. Self	2. Husband
3. Wife	4. Father
5. Mother	6. Other (specify):

3.4 Marital status of the household head

1. Married	2. Living together
3. Widowed	4. Divorced
5. Separated	6. Single, unmarried

3.5 Occupation of the household head

1. Professional: teacher, nurse, accounts, administrator, etc.	2. Clerical / secretarial
3. Trader / businessman / driver with own car, etc.	4. Employed tradesman, driver without own car, builder, etc.
5. Farmer / labourer / domestic worker	6. Other (specify)
8. NK	9. NA

4.0 ECONOMIC CHARACTERISTICS OF HOUSEHOLD

4.1 Does anyone in the household own their own farm?

1. Yes	2. No

4.2 What do they grow?

1. Food items, mainly for home consumption	2. Food items, mainly for sale on the market
9. NA	

4.3 Does anyone in the household own a cash crop?

1. Yes	3. No

4.4 If yes, what is the average size of cash crop farm owned by all the household members?

1. Less than 10 acres	2. 10 to 19 acres
3. 20 to 29 acres	4. 30 to 39 acres
5. 40 to 49	6. 50 to 59
7. 60 to 69	8. 70 to 79
9. 80 to 89	10. 90 to 99
11. 100+ acres	12. NA

4.5 What kind of toilet facility does your household use?

1. Flush latrine / WC
2. Ventilated improved pit / VIP / KVIP
3. Other pit latrine
4. Open fields
5. Defecate in house, feces transferred elsewhere / bucket latrine
6. Other (specify):

4.6 Does your household own or rent the house you live in?

1. Sole ownership	2. Joint ownership
3. Renting	4. Family / relation's house
5. House provided rent free	6. Perching
7. Other (specify):	8. NK

5.0 HEALTH HISTORY

5.1 What was your baby's birth weight? (kilograms) . **Check child's health record.** If the child's birth weight is no known, fill the boxes with "8888".

		.		

5.2 Has your child been previously hospitalized?

1. Yes	2. No

5.3 For what reason(s) was/were your child hospitalized?

1.	2.
3.	

5.4 Did your child use a bednet last night?	1. Yes	2. No

6.0 NUTRITION

6.1 How many times a day do you breastfeed?	1	2	3	4	5	6+
	9. NA (child is not breastfed)					

6.2 How old was your child when you first introduced foods other than breastmilk?		
	1. <4 months	2. 4-5 months
	3. 6 months	4. 7-8 months
	5. 9-10 months	6. >10 months
	8. NK	

6.3 In the last **7 days** has (*child's name*) had:

Cerelac……..	1. Yes	2. No
SMA……	1. Yes	2. No
Lactogen……..	1. Yes	2. No

6.4 What are the 3 most common foods that you add Sprinkles to?

1.	2.	3.

6.5 Ask the caregiver how she/he adds Sprinkles to food and describe below:

END OF DEMOGRAPHIC AND NUTRITION FORM, CHECK YOUR FORMS AND THANK THE RESPONDENT

Appendix C

Generic Model Specification

In general, the spatial model can be specified as:

$$\begin{aligned} Y_i \sim\ & f(\mu_i, \tau) \\ g(\mu_i) =\ & \beta X_i + \gamma W(s_i) + U(s_i) \\ \text{cov}\,[U(s+h), U(s)] =\ & \sigma^2 \rho(h, \phi) \end{aligned}$$

Where f was the distribution function for an observed value Y_i of observation i given it's expected value μ_i with a link function g and the τ variable containing additional parameters for f. The specific forms of the model used for each Aim were:

Aim 1: A log-Normal distribution f describing ferritin concentration, with τ variance parameter and exponential link function g;

Aim 2: A Binomial distribution f for prevalence of infection, with a logit link function g and τ unused; and

Aim 3: f Binomial and Poisson distributions with g logit and log link functions for prevalence and counts of infection, respectively. The τ parameter was again unused.

The transformed expected values $g(\mu_i)$ were determined by:

- X_i, a vector of individual-level explanatory variables (e.g. age, sex) for individual i;
- $W(s_i)$, a set of spatially referenced explanatory variables (e.g. elevation, distance to the nearest health facility) at location s_i for observation i;
- $U(s_i)$, a residual spatial effect allowing for variation in μ_i that is not explained by the covariates X_i and $W(s_i)$; and
- The regression coefficient parameters β and γ.

The residual spatial effect $U(s)$ was modelled as a stationary Gaussian random field with a Mat'ern correlation function[2]. For any two locations s_1 and s_2, the values $U(s_1)$ and $U(s_2)$ of the spatial surface were both Normally distributed with variance σ^2 and correlated with one another by an amount $\rho(s_1 - s_2, \phi)$, determined by a correlation function ρ and a range parameter ϕ. The correlation ρ decreased with distance, and thus when s_1 and s_2 were close to one another the values $U(s_1)$ and $U(s_2)$ tended to be more similar than when s_1 and s_2 were geographically farther apart. The range parameter ϕ controlled whether $U(s)$ changed quickly or gently in space. The range could also be interpreted as the distance (in km) above which correlations were less than 0.13, or $\rho(h, \phi) < 0.13$ when $|h| \geq \phi$. The Mat'ern correlation with a shape parameter fixed at 2 was used throughout, which produced 'gently rounded' surfaces that had two derivatives.

[2] Diggle P, Ribeiro PJ. Model-based geostatistics. New York, NY: Springer; 2007.

Appendix D

Baseline Comparisons of Participant Characteristics

Baseline characteristics of the Ghana trial participants at baseline, endline, and among those who were lost to follow-up

	Baseline	Endline	Lost to follow-up*
Trial participants with a geo-coded residence (n)	1943	1781	162
Males (%)	992 (51.1)	900 (50.5)	92 (56.8)
Enrolment age (months), mean (SD)	19.2 (8.5)	19.3 (8.5)	18.5 (8.6)
Serum ferritin (µg/L), geometric mean (SD)	35.1 (3.65)	34.95 (3.65)	37.02 (3.63)
CRP (mg/L), mean (SD)	3.34 (4.96)	3.32 (4.94)	3.56 (5.20)
Parasite density (count/µL), geometric mean (SD)	3003.0 (5.35)	2939.1 (5.34)	3784.7 (5.41)
Anthropometric status[a]			
Weight-for-length z-score, mean (SD)	-0.63 (0.97)	-0.62 (0.97)	-0.72 (0.96)
Length-for-age z-score, mean (SD)	-0.80 (1.29)	-0.80 (1.29)	-0.82 (1.35)
Maternal Education, n (%)[b]			
None	586 (33.5)	543 (33.0)	43 (41.4)
Primary	334 (19.1)	317 (19.2)	17 (16.4)
Middle	709 (40.5)	670 (40.7)	39 (37.5)
Secondary or higher	123 (7.02)	118 (7.16)	5 (4.81)
Household Asset Score, n (%)			

Low	866 (47.5)	817 (47.7)	49 (45.0)
High	957 (52.5)	897 (52.3)	60 (55.1)
Age of introduction of complementary foods, n (%)			
≤6 months	1573 (87.3)	1475 (87.0)	98 (91.6)
>6 months	229 (12.7)	220 (13.0)	9 (8.41)

*No significant differences between endline sample and those lost to follow up
[a]Estimated using the WHO Child Growth Standards
[b] Total n=1752 (74 respondents were not mothers, 117 missing due to incomplete surveys)

Appendix E

Supplementary Tables S5.1-S5.3

Table S5.1: Spatial model of individual-level factors associated with baseline infection status[1] among 1943 Ghanaian children (Brong-Ahafo Region, March-April 2010)

Covariates	Odds ratios (95% CrI)		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia					
Intercept	0.556	(0.366, 0.860)	3.831	0.608	0.008
Age per month			(1.875,	(0.407,	(0.031,
6-23 months	1.027	(1.007, 1.048)	6.973)	0.922)	0.004)
24-35 months	0.958	(0.923, 0.994)			
Sex (male reference)	1.061	(0.870, 1.292)			
Length-for-age z-score	0.898	(0.823, 0.980)			
Weight-for-length z-score	0.900	(0.811, 0.999)			
Baseline iron status	1.440	(1.321, 1.575)			
2) Inflammation without parasitaemia					

Intercept	0.105	(0.065, 0.167)	7.752	0.489	0.008
Age per month			(3.143,	(0.267,	(0.030,
6-23 months	0.990	(0.965, 1.016)	16.19)	0.906)	0.004)
24-35 months	0.959	(0.909, 1.010)			
Sex (male reference)	1.170	(0.901, 1.521)			
Length-for-age z-score	1.005	(0.895, 1.127)			
Weight-for-length z-score	0.900	(0.785, 1.032)			
Baseline iron status	1.183	(1.089, 1.284)			
3) Parasitaemia with fever					
Intercept	0.054	(0.023, 0.113)	5.046	1.001	0.008
Age per month			(1.756,	(0.649,	(0.030,
6-23 months	1.033	(0.996, 1.071)	13.30)	1.535)	0.004)
24-35 months	0.936	(0.870, 1.001)			
Sex (male reference)	1.129	(0.791, 1.612)			
Length-for-age z-score	0.957	(0.814, 1.121)			
Weight-for-length z-score	0.822	(0.676, 0.997)			
Baseline iron status	1.343	(1.223, 1.481)			
4) Parasitaemia					
Intercept	0.393	(0.217, 0.736)	3.334	1.019	0.008
Age per month			(2.005,	(0.710,	(0.033,
6-23 months	1.051	(1.025, 1.077)	5.371)	1.482)	0.004)
24-35 months	0.977	(0.937, 1.018)			
Sex (male reference)	0.963	(0.762, 1.217)			

Length-for-age z-score	0.865	(0.778, 0.960)
Weight-for-length z-score	0.920	(0.811, 1.042)
Baseline iron status	1.274	(1.175, 1.386)

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (binary): 1 = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours), 0 = any malaria parasitaemia without concurrent fever or history of reported fever;

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

Baseline iron status = iron status at baseline, defined as serum ferritin concentration corrected for CRP using the regression method (Namaste et al.)

Table S5.2: Spatial model of household-level factors associated with infection status[1] among 1943 Ghanaian children (Brong-Ahafo Region, March-April 2010)

Covariates	Odds ratios (95% CrI)		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia					
Intercept	0.541	(0.322, 0.866)	4.162	0.538	0.007
Asset score	1.050	(0.938, 1.174)	(1.828,	(0.346,	(0.028,
Maternal education	0.752	(0.603, 0.939)	8.227)	0.866)	0.004)
Distance to health facility (km)	1.152	(1.055, 1.277)			
2) Inflammation without parasitaemia					
Intercept	0.183	(0.115, 0.285)	7.277	0.416	0.008
Asset score	0.991	(0.860, 1.141)	(2.661,	(0.221,	(0.029,
Maternal education	1.107	(0.821, 1.504)	16.26)	0.795)	0.004)
Distance to health facility (km)	0.918	(0.839, 1.001)			
3) Parasitaemia with fever					
Intercept	0.033	(0.013, 0.071)	8.068	0.867	0.008
Asset score	1.079	(0.868, 1.337)	(3.180,	(0.460,	(0.029,
Maternal education	0.893	(0.602, 1.337)	16.33)	1.640)	0.004)
Distance to health facility (km)	1.241	(1.076, 1.458)			

4) Parasitaemia					
Intercept	0.228	(0.115, 0.420)	3.795	0.721	0.008
Asset score	1.085	(0.942, 1.249)	(1.604,	(0.479,	(0.031,
Maternal education	0.638	(0.494, 0.827)	7.813)	1.101)	0.004)
Distance to health facility (km)	1.248	(1.113, 1.425)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (binary): 1 = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours), 0 = any malaria parasitaemia without concurrent fever or history of reported fever;

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

Table S5.3: Spatial model of satellite-derived geo-spatial factors associated with baseline infection status[1] among 1943 Ghanaian children (Brong-Ahafo Region, March-April 2010)

Covariates	Odds ratios (95% CrI)		Range parameter in km (95% CrI)	Standard deviations of random effects (95% CrI)	
				Spatial	Compound
1) Inflammation and/or parasitaemia					
Intercept	0.561	(0.298, 1.017)	5.827	0.637	0.008
Elevation (m)	0.991	(0.985, 0.997)	(2.844,	(0.398,	(0.030,
Urban/built up land (LC13)	0.810	(0.466, 1.387)	10.42)	1.043)	0.004)
Woody savannahs (LC8)	0.744	(0.261, 2.081)			
NDVI	0.722	(0.151, 3.318)			
NDVI*LC8	0.683	(0.263, 1.784)			
NDVI*LC13	3.095	(0.696, 14.88)			
2) Inflammation without parasitaemia					
Intercept	0.117	(0.063, 0.215)	6.663	0.486	0.008
Elevation (m)	1.001	(0.995, 1.006)	(2.706,	(0.260,	(0.029,
Urban/built up land (LC13)	1.135	(0.614, 2.067)	13.84)	0.915)	0.004)
Woody savannahs (LC8)	0.865	(0.145, 3.424)			
NDVI	1.019	(0.183, 5.401)			
NDVI*LC8	0.805	(0.272, 2.637)			

NDVI*LC13	1.436	(0.299, 7.321)			
3) Parasitaemia with fever					
Intercept	0.043	(0.013, 0.112)	9.251	1.062	0.008
Elevation (m)	0.990	(0.981, 0.999)	(3.233,	(0.584,	(0.030,
Urban/built up land (LC13)	0.757	(0.213, 2.399)	19.85)	1.985)	0.004)
Woody savannahs (LC8)	1.268	(0.192, 5.738)			
NDVI	0.485	(0.020, 8.294)			
NDVI*LC8	0.729	(0.179, 3.308)			
NDVI*LC13	4.835	(0.255, 155.3)			
4) Parasitaemia					
Intercept	0.360	(0.143, 0.856)	4.396	0.982	0.008
Elevation (m)	0.990	(0.981, 0.999)	(2.034,	(0.643,	(0.032,
Urban/built up land (LC13)	0.488	(0.184, 1.192)	8.847)	1.498)	0.004)
Woody savannahs (LC8)	0.688	(0.204, 2.197)			
NDVI	0.325	(0.022, 4.002)			
NDVI*LC8	0.828	(0.261, 2.711)			
NDVI*LC13	4.884	(0.349, 88.71)			

[1]Infection status definitions:

1) Inflammation and/or parasitaemia (binary): 1 = CRP > 5 mg/L and/or any malaria parasitaemia, 0 = CRP ≤ 5 mg/L and absence of parasitaemia;

2) Inflammation without parasitaemia (binary): 1 = CRP > 5 mg/L without malaria parasitaemia, 0 = CRP ≤ 5 mg/L without parasitaemia;

3) Parasitaemia with fever (binary): 1 = any malaria parasitaemia with concurrent fever (axillary temperature >37.5^{0}C) or history of reported fever (within 48 hours), 0 = any malaria parasitaemia without concurrent fever or history of reported fever;

4) Parasitaemia (binary): 1 = any malaria parasitaemia (with/without fever), 0 = absence of parasitaemia with/without fever

Model prior shape = 1.117, model prior rate = 0.157

CrI = credible interval

NDVI = normalized difference vegetation index, averaged over 2010 yearly values, centered by dividing by 1000 and subtracting 4