Social and Environmental Epidemiology of Schistosomiasis in Ghana

Irene Akosua Larbi, M.Sc (Med)

Submitted for the degree of Doctor of Philosophy

at Lancaster University,

October, 2015.

Abstract

Chemotherapy has provided a realistic approach for controlling schistosomiasis in resource-poor settings such as sub-Saharan Africa and control programmes have mainly adopted an age-targeted strategy of implementation. However, it is being increasingly argued that the setting-specific context in which the transmission of schistosome infections occurs may render this global approach of chemotherapy implementation inefficient. Evidence from different endemic settings points to the fact that the transmission dynamics of schistosome infections is not merely an interplay between humans and the parasites, but also a series complex interactions between environmental and social processes. Hence the degree of the spatial heterogeneity, that often characterises the transmission of infections, may differ for different endemic settings depending on the extent of the interaction between these processes. This thesis employs geostatistical methodology in assessing the collective effects of the social and environmental determinants of schistosome transmission within different endemic settings in Ghana. It also explores how these processes may influence the patterns of transmission at the local level and how these patterns could be utilised in improving the effectiveness of mass chemotherapy intervention programmes.

Acknowledgements

I first of all like to acknowledge the Economic and Social Research Council (ESRC) for the three-year grant that brought me to Lancaster; and Professors Daniel Boakye and Maria Yazdanbakhsh for granting me access to the datasets I used for this thesis. I am also deeply indebted to the Faculty's Postgraduate Programmes' Administrator, Ms. Dawn McCracken, for her constant support throughout those anni horribiles.

I appreciate the kind support of Carmen Armero and her wonderful textbook recommendations; Benjamin Taylor for his critical review of the paper I submitted for my upgrade; Barry Rowlingson for coming to my rescue whenever I got stuck with certain aspects of the work; Emanuele Giorgi for taking the time to demonstrate the functioning of his *PrevMap* package; Michelle Stanton and Nicola Reeve for providing me with the necessary guidance during those early days when I had to learn `R` and La*T*EX from scratch; and my dearest friends Ebrahim Al Hanaei, Mr & Mrs Gerald Brady and Mr & Mrs Appiah for their unwavering support.

When I think of the curious paradox of the emotional turmoil I experienced while studying on this programme, the opening paragraph of Charles Dickens' novel, *A Tale of Two Cities*, springs to mind: *"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness...it was the spring of hope, it was the winter of despair"*. That pretty much captures most of it but to that I might add that it was also a period of much enthusiasm and apathy, trust and mistrust, anxiety, self-realisation and absolute misery. Then of course, there was the constant sense of alienation and the consequential social isolation. At the end of it all, the adverse effects including the lost time and the emotional adjustment make one wonder whether it was all worth it. I really hope it was. In any case, I am awfully happy to see the end of this burdensome, nightmarish, panic-inducing and sleep-depriving chapter of my life. Now, on to the next.

Dedication

Declaration

This thesis is entirely my own work and has not been submitted in full or in part for the award of a higher degree at any other educational institution. No sections of this thesis have been published.

Irene Akosua Larbi, October 2015

Contents

Contents

List of Figures

List of Tables

Part 1

List of Figures

List of Tables

Part II

List of Figures

List of Tables

Part III

List of Figures

List of Tables

1 General Introduction

1.1 Epidemiology of Schistosomiasis in Ghana

Ghana is classified by the World Health Organisation as one of the ten most endemic countries for schistosomiasis in the world. These highly endemic countries, which are all located in Africa, are collectively estimated to account for about 67% of the global population in endemic areas that requires treatment (World Health Organization 2013). The country is endemic for two forms of human schistosomiasis: the urinary form of the disease caused by *Schistosoma haematobium* and the intestinal form caused by *S. mansoni* (World Health Organization 2012). However, while both forms of the disease are distributed nationwide, *S. haematobium* tends to be more common with over 80% of the total area of the country being classified as meso-endemic areas. By contrast, about 90% of the country are low endemicity areas for *S. mansoni* infection (Figures 1 and 2).

While the endemicity of schistosomiasis in Ghana may be directly linked to the post-independence water resource development projects that occurred across the country mainly in the 1960s and 70s, evidence from studies pre-dating these projects suggests schistosomiasis already existed (Mahmoud 2001, Hunter 2003, Hunter et al. 1982). Therefore, the construction of the water impoundments projects in the endemic zones of schistosomiasis increased its transmission foci and triggered off epidemics in the surrounding areas of these projects. For instance, an 80% increase in the prevalence of *S. haematobium* infection was recorded in children in the shoreline communities of the Volta Lake within a year of reaching its maximum level in 1969 (Hunter et al. 1982, Mahmoud 2001, Hunter 2003).

1.2 Preventive Chemotherapy Interventions in Ghana

Due to the nationwide distribution of schistosomiasis in Ghana, the entire population is considered to be at risk. Consequently, provisional estimates by the World Health Organisation in 2010 suggested that the nation's 24,332,755 population required preventive chemotherapy for schistosomiasis (World Health Organization 2012). However, in keeping with the age-targeted global implementation strategy where interventions

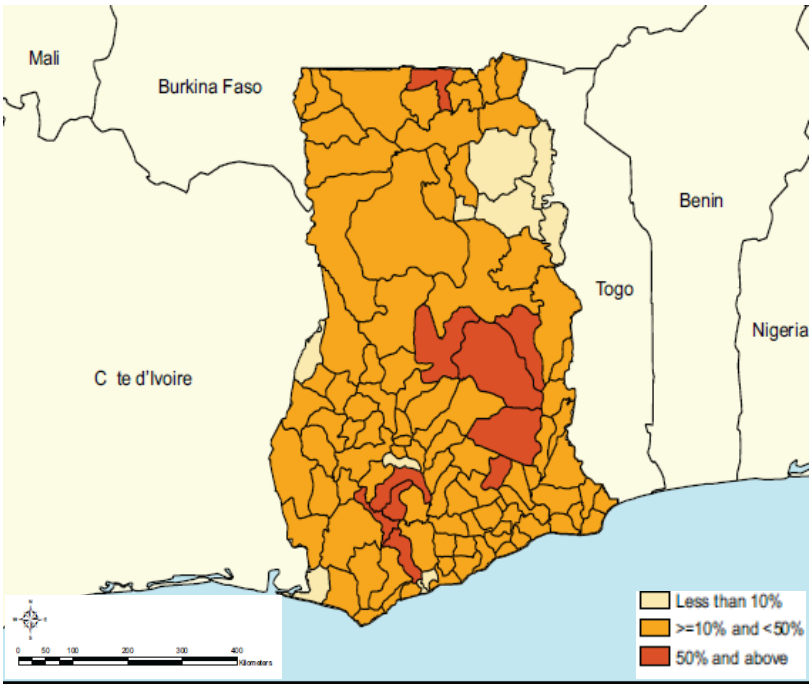


Figure 1: Map of Ghana showing areas with low, moderate and high endemicity levels of *S. haematobium* infection in 2010 (adapted from World Health Organization 2012)

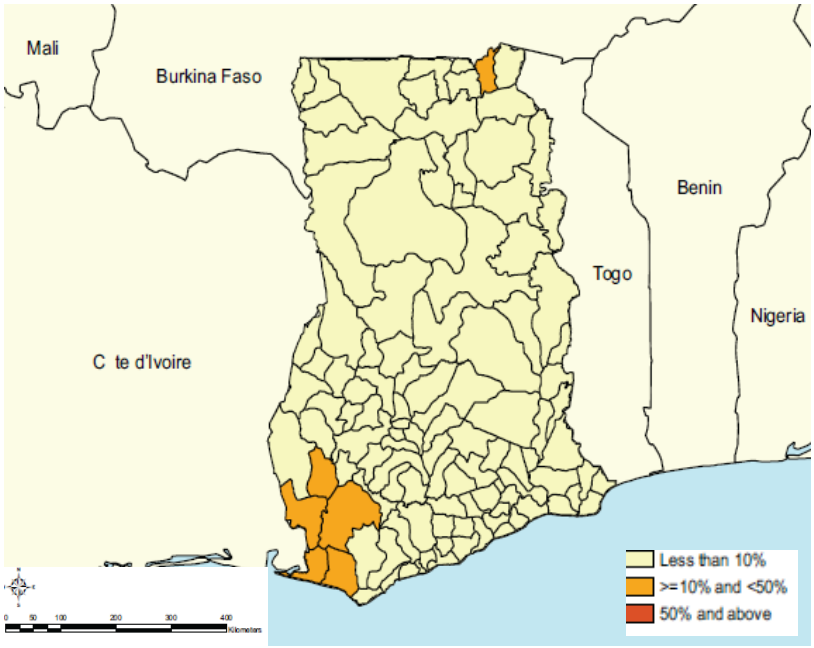


Figure 2: Map of Ghana showing areas with low and moderate endemicity levels of *S. mansoni* infection in 2010 (adapted from World Health Organization 2012)

are primarily focused on groups that are most susceptible to acquiring heavy infections and experiencing the greatest morbidity (Basanez et al. 2012); ongoing chemotherapy intervention programmes across the country have mainly been school-based with children in the school-age population (6-15 years) as the primary target. Therefore, the national coverage of 27.5% that was attained in 2010 was exclusive to school-age children, who constituted 24% (5,866,905) of the country's population (Figure 3) (World Health Organization 2013, 2012).

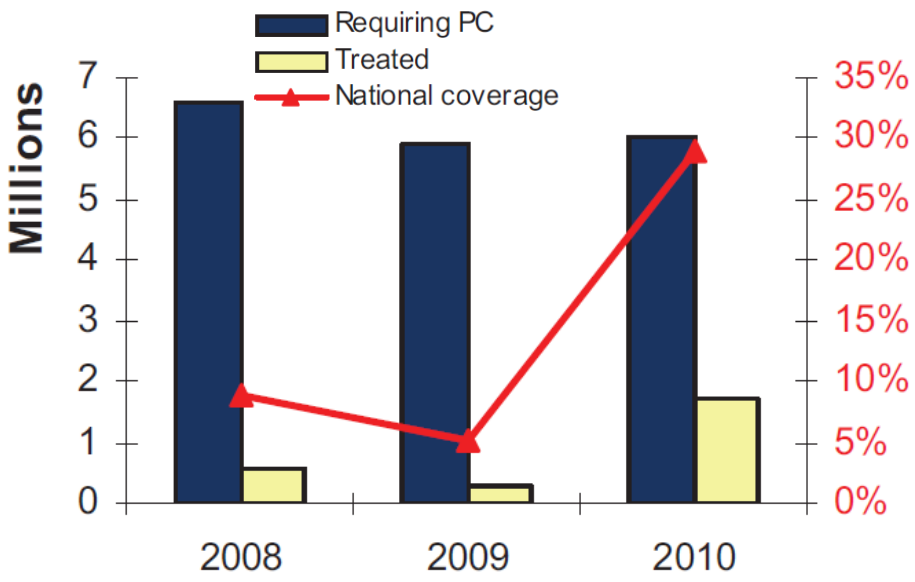


Figure 3: National coverage of preventive chemotherapy intervention for schistosomiasis in Ghana as of 2010 (Adapted from World Health Organization 2012).

However, long before these recent national chemotherapy intervention programmes were ever implemented, pilot projects for a nationwide control programme were conducted as far back as the mid 1970's. These projects were necessitated by the explosive increase in the transmission rates of *S. haematobium* following the creation of the Volta Lake in 1964 (Hunter et al. 1982, Hunter & Organization 1993). Therefore, the pilot projects explored different control strategies, including selective chemotherapy with metrifonate, focal molluscicide of water contact sites and health education; over a 37 mile stretch of the Lake's 3,106 mile shoreline (Mahmoud 2001, Hunter & Organization 1993, Chu et al. 1981). However, despite the success of these early trials and their subsequent adoption into the primary health care system at the time, they were never implemented at the national level (Mahmoud 2001).

1.3 Life-cycle of the Human Schistosomes

Human schistosomiasis is caused by parasitic blood-dwelling trematodes of the genus *Schistosoma*. The schistosomes are dioecious digeneans whose life-cycle consists of two alternating phases: a sexual phase that occurs within the human definitive host; and an asexual phase which occurs within specific species of freshwater molluscs that constitute the intermediate hosts (Figure 4) (Colley et al. 2014). The African species of schistosomes, *Schistosoma haematobium* and *S. mansoni*, are transmitted by pulmonate snails of the genus *Bulinus* and *Biomphalaria* (which are classified under the class *Gastropoda*, sub-class *Pulmonata* and family *Planorbidae*) (Mahmoud 2001).

1.3.1 Sexual Phase

The adult worms, which measure between six and 26 millimetres in length, either inhabit the perivesicle venous plexus or the mesenteric venules of the human host, depending on the species (Mahmoud 2001). Though their average lifespan generally ranges between three to 10 years, the schistosomes may survive for up to 40 years in some cases (Colley et al. 2014). The female worms are enclosed within the gynaecophoric canals of the males where they typically produce hundreds of eggs on daily basis throughout life (Figure 4 E). The eggs, which measure between 60-400μm each in length, are fertilised by the male worms upon shedding (Colley et al. 2014, Mahmoud 2001).

Depending on the species, these eggs may either migrate to lumen of the bladder or the intestine to be excreted in urine or faeces within 7-10 days of oviposition. During this perivesicle or peri-intestinal migration, the eggs mature fully into their embryonated forms. Though the shed eggs are intended to be excreted from the body of the host, at least from the perspective of the worms, a fraction always tends to get trapped in the body tissues during their migration to the exterior (Colley & Secor 2014). According to estimates by early studies, the egg retention rate is approximately one egg per worm pair per day (Jordan & Webbe 1982). Subsequent studies went to suggest the egg retention rate may be as high as 50% (Medley & Bundy 1996). These tissue-trapped eggs, which die within 2-3 weeks, are responsible for inducing the inflammatory reactions which ultimately cause the pathological effects of schistosomiasis (Mahmoud 2001).

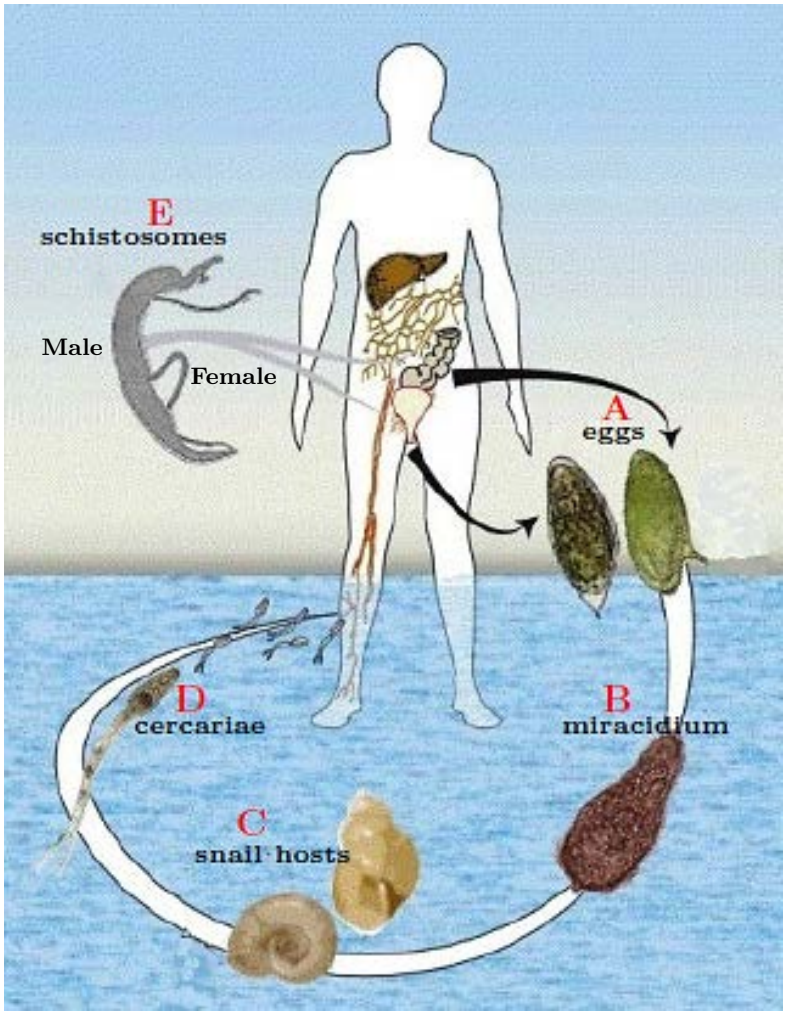


Figure 4: Life-cycle of the human schistosomes showing the two alternating phases in the human definitive and the intermediate freshwater molluscan hosts (adapted from Gryseels et al. 2006 with permission of the rights holder, Elsevier)

1.3.2 Asexual Phase

The excreted schistosome eggs may remain viable for seven days (Gryseels et al. 2006). Upon contact with freshwater, the eggs hatch under suitable conditions, such as a temperature range of $\geq$10 - 30° Celsius and light intensity, to release the free-living ciliated miracidia measuring 160μm by 65μm in size (Figure 4B). The miracidia then go through what some authors have termed the "scanning phase" where they randomly swim about until they locate and infect a suitable intermediate snail host, guided by stimuli. The location of the snail hosts is, however, influenced by conditions of the water body including the turbulence, velocity of flow and temperature. The miracidia generally remain infective in the freshwater body for 8 - 12 hours (Mahmoud 2001, Jordan & Webbe 1982).

After infecting the snail hosts, the miracidia undergo two phases of asexual replication: firstly, into the multicellular primary and subsequently, secondary sporocysts. In the second phase, thousands of cercariae are produced from the secondary sporocysts and shed into the freshwater body in a pattern that follows periodic peaks of output. This pattern of cercarial shedding is mainly attributed to a species-specific innate rhythm of the snail hosts. Both the miracidia and cercariae are non-feeding stages that rely on glycogen energy reserves. Therefore, their infectivity is largely influenced by the period it takes for their energy reserve to deplete. In all, the asexual phase of the life-cycle takes between 4-6 weeks to complete (Mahmoud 2001, Colley et al. 2014, Jordan & Webbe 1982).

1.3.3 Exposure-Infection Disease Processes

Each miracidium infection in the intermediate snail host results in the production and the subsequent shedding of thousands of same-sex cercariae. The cercariae are mainly shed during daytime and tend to survive for 72 hours during which time they must locate and penetrate the skin of the human host. Infection is, therefore, initiated when the cercariae burrow into the epidermis of the exposed human host (Figure 4). After transforming into young worms known as the schistosomula, they migrate in the systemic venous circulation of the human host to the hepatic sinusoids via the lungs and mature into adult schistosomes (Gryseels et al. 2006, Mahmoud 2001).

If the human host were assumed to have an equally likely chance of being infected by both male and female cercariae, then any constantly exposed host would ideally end up with worms of both sexes. These adult male and female worms are generally thought to undergo monogamous pairing before each worm pair copulates and migrate to their final destination, the perivesicle venous plexus (in the case of *S. haematobium* infection) or the mesenteric venules (in the case *S. mansoni* infection). On reaching their final destination, it is also generally assumed that each mated worm pair would undergo oviposition with the female laying eggs continuously throughout life. The development of the cercariae to the egg-producing adult stage takes a period of 5 - 7 weeks (Colley et al. 2014, Mahmoud 2001).

1.3.3.1 Egg-Patency Versus Host-Antibody Patency

A T-helper 1 (T_H1) immune profile, which has a protective effect, is initially induced in response to the antigens from the maturing worms in the portal systems. Hence, the pre-patent period, i.e. 3 - 5 weeks after the initiation of infection, is mainly characterised by an elevated production of T_H1 - associated cytokines, including interleukins γ and 12 as well as tumor-necrosis factor-α (TNF-α). While the host immune system tends to tolerate the live adult worms without producing any inflammatory response, their death is thought to provoke the production of immunogens. These immunogens are associated with protective immune responses directed at the antigens of the migrating schistosomula (Mahmoud 2001, Pearce & MacDonald 2002, Colley et al. 2014).

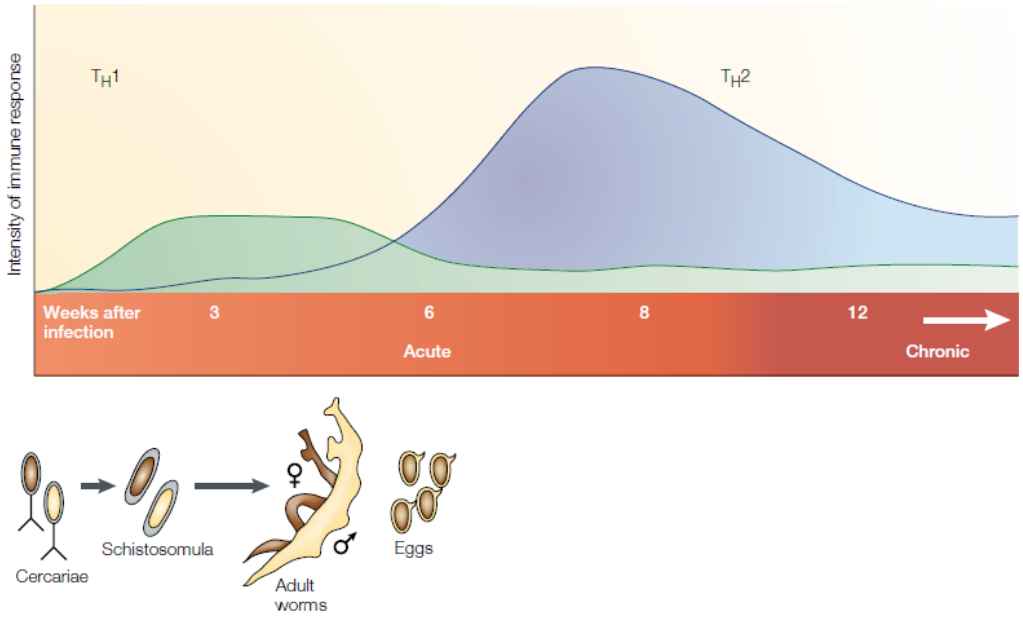


Figure 5: Host immune responses associated with schistosome infection (adapted from Pearce & MacDonald 2002 with permission of the rights holder, Nature Publishing Group).

At the onset of egg deposition, which occurs within 5-6 weeks after the initiation of infection, a T_H2 immune profile is triggered by specific egg antigens. The emergence of the T_H2 profile together with its associated cytokines (interleukins 4 and 10; and also elevated TNF-α), modulates the existing T_H1 response from the pre-patent phase of infection (Figure 5). This step is deemed necessary for the granulomatous response to the schistosome eggs (Mahmoud 2001, Colley & Secor 2014).

1.3.3.1.1 Granulomatous Response

Matured miracidia-containing schistosome eggs are known to secrete antigenic histolytic substances and other proteases through the micropores that perforate their shells (Mahmoud 2001). These proteases, which facilitate the migration of the eggs by lysing the surrounding tissues, are also thought to have a potentially toxic effect on the host tissues. Therefore, when trapped in the body tissues, the secretions from the eggs may cause necrosis. As a defence mechanism against the effect of the egg antigens, the host forms granulomas around the eggs to sequester the secretions (Figure 6) (Colley & Secor 2014, Pearce & MacDonald 2002). This granuloma reaction is thought to be most intense during the first 8-10 weeks but diminishes as the T_H2 response becomes modulated over time. Therefore, during advanced stages of infection, the tissue-trapped eggs get destroyed at a faster rate due to their smaller granulomas.

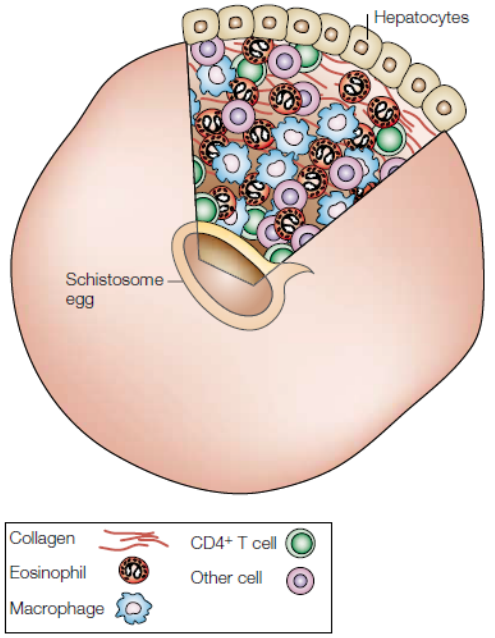


Figure 6: A schematic diagram of a granuloma around a tissue-trapped ovum (adapted from Pearce & MacDonald 2002).

Upon the death of the eggs, the granulomas resolve into deposits of fibrotic plaques which build-up progressively over time and ultimately become responsible for inducing the pathological effects of schistosomiasis. Hence, common pathologies such as increased portal blood pressure, ascites and portal-systemic venous shunts that occur during *S. mansoni* infection are the results of the accumulation of fibrotic plaques in the body. *S.*

haematobium infection may also result in the development of bladder cancer (Pearce & MacDonald 2002).

It has been suggested that in high transmission areas where heavy intensity infections prevail, the accumulation of eggs in the body tissues may surpass the rate of their destruction by the granulomas. In such instances, the granulomas tend to accumulate over time and become calcified in the process. Consequently, both previous and current infections may become responsible for the pathological effects of schistosomiasis at any given point in time (Medley & Bundy 1996).

1.4 Snail Hosts Species in Ghana

Ghana is endemic for two forms of human schistosomiasis caused by *Schistosoma haematobium* and *S. mansoni*. Each of these schistosome species is transmitted by its own specific range of freshwater molluscs. Therefore, the distribution of these schistosomes across the country is influenced by the habitat range of their respective intermediate molluscan hosts (Colley et al. 2014). Two species of intermediate hosts, *Bulinus truncatus (rohlfsi)* and *B. globosus*, are responsible for transmitting *S. haematobium* in the country. These species are regarded as the primary and secondary hosts of *S. haematobium*, respectively. Studies have shown that the strain of *S. haematobium* transmitted naturally by either molluscan host tends to exhibit low infectivity for the other host species (Figure 7). The intermediate host of *S. mansoni*, on the other hand, is *Biomphalaria pfeifferi* (Brown 2002).

1.4.1 Conditions of the Aquatic Environment

The intermediate molluscan hosts of the schistosomes generally tend to thrive in the shallow margins of low-velocity freshwater bodies including natural streams, ponds and impounded waters (Mahmoud 2001). The suitability of these freshwater habitats is, however, determined by a set of specific conditions. Prominent among these are: temperature, salinity, pH and ecological changes resulting from water source modification.

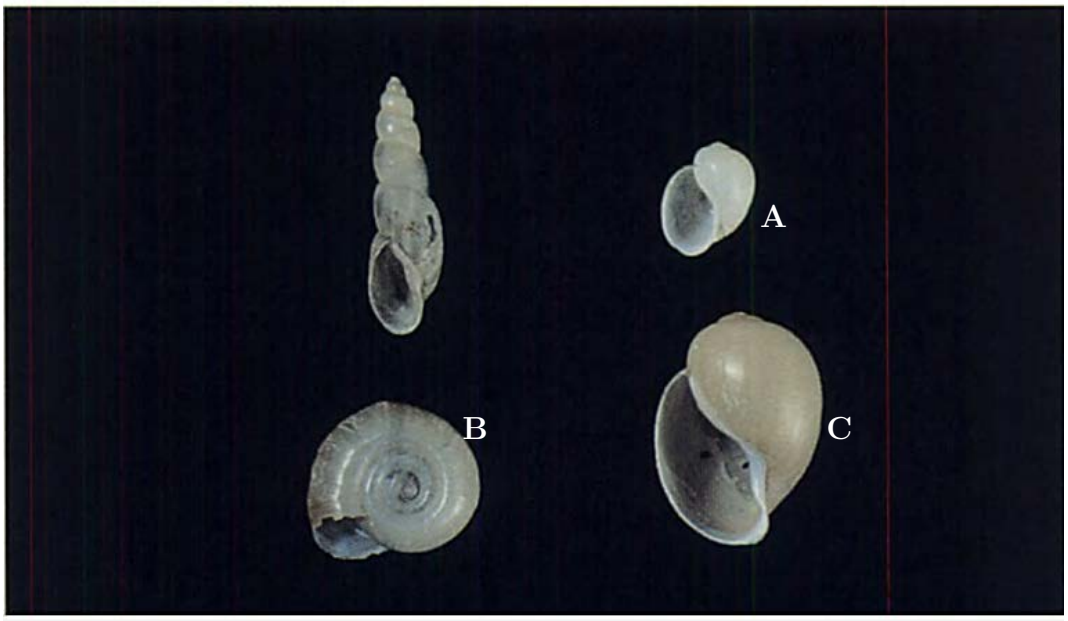


Figure 7: The intermediate molluscan hosts of the African schistosomes in Ghana: *Bulinus truncatus* (A), *Biomphalaria pfeifferi* (B), *Bulinus globosus* (C) (adapted from Mahmoud 2001).

1.4.1.1 Temperature

Studies have shown that the survival and development of the intermediate molluscan hosts within their freshwater habitats are favoured by a temperature range of 25 ± 2° Celsius. This optimal temperature ensures their breeding, egg incubation, growth and the completion of their annual cycles (Mahmoud 2001, Brown 2002). However, the occurrence of these molluscs in temperatures below the optimum is also possible. Indeed studies in endemic areas have shown that the intermediate molluscs may also thrive in higher altitudes where the mean temperatures are often too low to favour the development of the schistosome parasites. Therefore, snails that occur in such lower temperatures do not transmit any schistosome infections. Temperatures above 30° Celsius are, however, known to cause thermal death in these molluscs (Brooker 2007).

1.4.1.2 Salinity

Experimental studies have indicated that varying levels of salinity may also influence the growth and fecundity of the intermediate molluscan hosts of the schistosomes. Studies on *Biomphalaria pfeifferi* found conductivities of 350 - 500 micromhos to be optimum

for their survival, fecundity and growth (Brown 2002). Therefore, this species rarely occurs in estuarine conditions where salinity levels may be above their tolerance range (Mahmoud 2001). However, *Bulinus* spp tends to tolerate higher conductivities than *Biomphalaria pfeifferi* though salinity levels of up to 2000 micromhos, which characterises estuarine conditions, may also prove deleterious for their survival (Brown 2002, Mahmoud 2001). In Ghana, the transmission of *S. haematobium* by *Bulinus rohlfsi* is known to occur in the Ke Lagoon, thereby suggesting the suitability of the salinity levels of that lagoon for the survival of *Bulinus rohlfsi* (Brown 2002).

1.4.1.3 pH

Acidic conditions have proved uncongenial for the intermediate molluscan hosts through their tendency to induce the snails mucus to coagulate. Such conditions are also known to have a damaging effect the snails' shells. Therefore, the favourable habitats of the molluscan hosts are generally restricted to alkaline conditions (Brown 2002).

1.4.1.4 Water Resource Development

The construction of the Akosombo Dam over River Volta in 1964 resulted in the creation of the Volta Lake, which is the largest man-made lake in the world as judged by its surface area. The impact of this impoundment, including the change in velocity of the water as well as the seasonal fluctuations in the water level proved favourable for the survival and distribution of the population of *Bulinus truncatus rohlfsi* that already occupied the Volta basin. Therefore, the explosive increase in the prevalence of *S. haematobium* that occurred in the shoreline communities of the new Lake was primarily attributable to an increase in the population density of *Bulinus truncatus rohlfsi*, which serves as an intermediate host for *S. haematobium* (Mahmoud 2001, Hunter & Organization 1993).

In the early stages of water impoundments, the aquatic ecosystem is usually modified by the establishment of populations of submerged aquatic flora, including *Ceratophyllum* spp, *Pistia* and *Salvinia* which serve as suitable habitats for the intermediate hosts of the schistosomes. In the Volta Lake, studies have shown that the success of *Bulinus truncatus rohlfsi* is primarily attributable to populations of *Ceratophyllum* spp which

provide habitats and protection for the snails and their eggs. Moreover, other submerged vegetation such as *Spirodela* play an essential role in the distribution of the snails whilst *Polygonum*, *Alternanthera* and *Jussiaea* which characterise the Lake's margin and survive flooding ensure the establishment of these snails (Hunter & Organization 1993, Brown 2002). Studies have also shown that the annual fluctuations in the level of the Volta Lake result in the colonisation of exposed slopes by new vegetation which subsequently add on to the snails' biotopes. Moreover, the stability of the water level during the annual draw-down of the Lake in November - February provides conditions suitable for the expansion of the population of *Bulinus truncatus (rohlfsi)* (Brown 2002).

1.5 Diagnostic Assessments of Schistosomiasis

There is as yet no gold standard technique for diagnosing human schistosomiasis. The available detection techniques include direct parasitological detection, indirect assessment of pathological and clinical markers of the disease as well as immuno-diagnostic techniques (Table 1).

1.5.1 Parasitological Detection

The direct detection of viable schistosome eggs in the excreta of the human host, using parasitological techniques, has been the diagnostic standard for active schistosomiasis (Colley et al. 2014). These techniques involve the use of urine sedimentation or filtration methods for the detection of *S. haematobium* eggs; and the Kato-Katz technique for *S. mansoni* eggs (Mahmoud 2001). Due the distinctive shape and size of the eggs of the different species of schistosomes, this method of detection is 100 percent specific (Stothard 2009). The operational cost associated with this method is also low and it is suitable for use under field conditions where the laboratorial structure is often improvised (Rabello 1997). Therefore, the method has mainly proved useful for epidemiological surveys and the crude mapping of infections for field-based schistosomiasis control programmes (Colley et al. 2014, Cavalcanti et al. 2013).

Despite its efficacy and cost-effectiveness, however, the sensitivity of parasitological detection is often confounded by specific aspects of the biology of the schistosomes, including the sporadic egg excretion or reversion of infection that mainly characterises

Table 1: Detection methods for human schistosomiasis, with emphasis on *S. haematobium* and *S. mansoni* infections (adapted from Feldmeier & Poggensee 1993 with permission of the rights holder, Elsevier).

Category	Method	Detection of	Applicable for infection with			
			S. haem.	*S. mans.*	*S. jap.*[a]	*S. int.*
Direct; parasitological	examination of urine, stool, rectal mucosa	eggs	+	+	+	+
Indirect; assessment of specific pathology through clinical, biochemical and immunological disease markers	looking/asking for symptoms and signs	gross haematuria	+			
		bloody diarrhea		+	(+)	+
		microhaematuria,	+			(+)
		proteinuria,	+	(+)		
		leukocyturia,	+			
		occult blood in stool		+	(+)	+
	looking for organ involvement, assessment of altered immune functions	sonographical alterations;[b] altered liver haemodynamics and	+	+	+	(−)
		collagen synthesis, altered cell-mediated		+	(+)	(−)
		immunity	+	+	+	(−)
Indirect; direct; immunological	serology	specific antibodies	+	+	+	+
		schistosome-derived antigens	+	+	+	+
	in vivo immediate and delayed-type hypersensitivity	previous sensitization	+	+	+	(−)

low intensity infections (Stothard 2009, Jordan & Webbe 1982). Hence, there tend to be a direct proportionality between the sensitivity of parasitological detection and the intensity of infection, thereby rendering it unsuitable for use in low endemicity settings (Cavalcanti et al. 2013). The sample size and the level of experience of the technician performing the examination may also affect the sensitivity of this method of detection (Mahmoud 2001).

Moreover, the schistosome egg excretion rate in the infected host generally exhibits diurnal well as day-to-day variation in output (Mahmoud 2001, Jordan & Webbe 1982). Therefore, the accuracy of this method has relied heavily on the examination of multiple samples per individual, which is often logistically infeasible especially for stool samples. Parasitological detection methods may also prove unsuitable for diagnosing atypical forms of the disease such as neuro-schistosomiasis (Cavalcanti et al. 2013).

1.5.2 Indirect Clinical Markers

Visible and cryptic blood in urine, i.e. macro and micro-haematuria, which normally characterise egg excretion during the early stages of *S. haematobium* infection are commonly used as indirect markers for the disease. Haematuria occurs as a result of the damage caused by schistosome eggs as they traverse the bladder wall during *S. haematobium* infection (Mahmoud 2001). Studies have shown that about 41 - 100 percent of the population of constantly-exposed infected children in pre-intervention areas normally test positive for micro-haematuria whilst 0-97 percent often show signs of gross haematuria (Gryseels et al. 2006). Therefore, gross haematuria has been utilised in school-based questionnaires for the rapid assessment of *S. haematobium* infection in high risk areas (Lengeler et al. 2002). While faecal occult blood tests may also serve as markers of *S. mansoni* infection, their sensitivity is hampered by the occlusion of blood (resulting from the perforation of the bowel by schistosome eggs) (Stothard 2009).

1.5.3 Immuno-Diagnostic Techniques

In bid to overcome the limitations associated with direct parasitological detection, immunodiagnostic techniques have mainly focused on the detection of the parasites' antigenic

components in serum and urine samples. This approach consequently led to the development of the genus cross-specific mono-clonal antibody-based enzyme-linked immunosorbent assays (ELISA) which measured the levels of schistosome gut-derived circulating antigens in serum. Later on, the technique was extended to the use of urine samples. However, on its evaluation, the technique was found to be biased towards the detection of *S. mansoni* infection while having a rather poor sensitivity for *S. haematobium* (Stothard 2009, Corstjens et al. 2008).

Subsequent diagnostic techniques in this area, therefore, focused more on the detection on *S. mansoni* infection. Thus far, the point-of-care (POC) lateral flow cassette assay is the most important of such techniques. The POC assay detects *S. mansoni* infection by measuring schistosome circulating cathodic antigens in urine samples. The evaluation of this technique in five sub-Saharan African countries, by the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE), has shown POC assays to be more sensitive, specific, rapid and less invasion than Kato-Katz (Colley et al. 2013). Moreover, while POC assays have a similar operational cost to Kato-Katz, they estimate the intensity of infection without being limited to the collected sample (Colley et al. 2013, Bergquist 2013).

1.5.4 Molecular Techniques

DNA-based detection methods, involving the use of multiplex real-time polymerase chain reaction (PCR), have also been developed. These techniques work by detecting schistosome DNA in faecal, urine and blood samples. However, despite their sensitivity, these techniques are also liable to sampling limitations when used on faecal samples (Colley et al. 2013, 2014).

1.6 Field-Collected Data

The data on which this thesis is based were specifically collected for two multi-centre immunological projects, the Innate Immune Responses & Immuno-Regulation in Schistosomiasis (SCHISTOINIR) and the Global View of Food Allergy, which had Ghana as one of the participating centres (Schistoinir 2012, GLOFAL 2006). The immunological

aspects of these projects were preceded by baseline cross-sectional parasitological and questionnaire surveys in the study sites. The studies presented in this thesis were, therefore, designed along the baseline data for the original projects that were made available for the thesis. Therefore, the present thesis is first of all limited by being based on data that were specifically collected for other purposes and hence did not necessarily fit the statistical and scientific concepts that were being explored. Moreover, the baseline surveys employed parasitological detection as an initial means of assessing the prevalence of schistosome infections in the targeted sites. Hence, inferences from this thesis may also be limited by the low sensitivity of the parasitological detection methods, especially for the low transmission post-intervention sites.

1.7 Organisation of Thesis

The main body of this thesis is organised into three parts:

Part I: Interventions for schistosomiasis have mainly focused on morbidity control through the age-targeted chemotherapy of high risk groups However, transmission control and the proper targeting of the susceptible parasite refugia within different endemic settings can only be attained when interventions are directed at entire population of infected individuals. This study, therefore, investigates how interactions between the macro and micro-level determinants of infectivity could be utilised in identifying groups of individuals with increased likelihoods of infection at the community-level.

Part II: The control of schistosomiasis relies primarily on mass chemotherapy with praziquantel. However, these chemotherapy interventions may also influence the epidemiology schistosomiasis in ways that ultimately increase the risk of resurgence. This study investigates the transmission patterns of persisting schistosome infections in low endemicity post-intervention areas, as a function of the socio-economic standards of households.

Part III: Interactions between the social and environmental determinants of schistosomiasis may occur in a dynamic state of change. This study investigates how such dynamic changes may ultimately influence transmission control in typical endemic settings where exposure to the sources of infection is virtually inevitable.

Part I

Spatial variation in the risk of *Schistosoma haematobium* and *S. mansoni* infections in rural endemic settings in southern Ghana

Abstract

Background The global control strategy for schistosomiasis relies on district-level data on school-aged children (6-15 years) in establishing modalities for the implementation of chemotherapy. However, the transmission of schistosomiasis is influenced by setting-specific factors which may in turn determine the degree of heterogeneity in the distribution of infected cases within any given endemic setting. Therefore, the same implementation strategy may not adequately control transmissions in all settings. However, the coverage and effectiveness of intervention programmes could be improved by the direct targeting of interventions at areas with increased likelihoods of infections within different settings. The present study, therefore, investigates some of the setting-specific factors that influence the local-level variation in the risk of *Schistosoma* infections as well as the heterogeneities in the transmission of infections within different endemic settings.

Materials and Methods Using point-referenced data from community-level cross-sectional surveys in three rural shoreline communities of the Volta Lake in Ghana, we examined the role of a series of setting-specific factors on the variation in the risk of schistosome infections. Our primary outcomes were the prevalence of *Schistosoma haematobium* and *S. mansoni* infection, detected by microscopy, in participants aged between five and 60 years. The probabilities of our study outcomes were modelled as functions of the measured covariates for each of these sites. As part of the models' validation, we conducted an assessment for residual spatial effects which would point to the effect of unobserved factors influencing the varying likelihoods in risk. A formal Monte Carlo test was used in assessing if any observed spatial trend may have occurred by chance.

Results Though the risk of our study outcomes were influenced by different factors within each of the studied communities, there were no obvious heterogeneities in the transmission of infections, as judged by a formal test for spatial trend in the residuals. The effects of the setting-specific micro-level determinants of infectivity, therefore, seemed negligible.

Conclusions The absence of significant residual spatial variation ruled out any significant heterogeneities in the distribution of infected cases across the studied shoreline communities. Our findings may be influence by the fact that though the micro-level determinants of infectivity may have acted at the group level, these groups of individuals may have necessarily lived in close proximities to each other.

1 Introduction

Schistosomiasis, a chronic parasitic infection caused by trematodes belonging to the genus *Schistosoma*, is regarded as the most important water-based disease globally (Steinmann et al. 2006, Brown 2011). It is mainly endemic in the rural parts of some tropical and sub-tropical regions of the world, but most of its current burden is concentrated in sub-Saharan Africa where active transmissions still occur in 46 countries (Steinmann et al. 2006). Though insidious in nature, chronic *Schistosoma* infections may ultimately result in associated morbidities, such as bladder and liver cancer, as well as developmental impairment pathologies that impact negatively on social and economic development in endemic areas (Lustigman et al. 2012, WHO 2005, Steinmann et al. 2006, Engels et al. 2002).

It is estimated that 280,000 deaths are attributable to schistosomiasis annually in sub-Saharan Africa (Fenwick et al. 2009). Moreover, schistosomiasis has been found to increase susceptibility to human immunodeficiency virus infection in women (Lustigman et al. 2012, WHO 2009). After malaria, schistosomiasis inflicts the highest disease burden in sub-Saharan Africa. Yet, it remains one of the neglected tropical diseases (Steinmann et al. 2006, Fenwick et al. 2009).

The transmission of schistosomiasis is focal and relies primarily on repeated exposure to lentic freshwater bodies that harbour the intermediate snail hosts of the schistosome species. As for other endemic parts of the world, the transmission of schistosomiasis in sub-Saharan Africa has been sustained over the years mainly by water resources development projects, such as dams and irrigation systems, that favour the ecology of the snail hosts (Steinmann et al. 2006). Moreover, as socio-economic development is low in most rural parts of sub-Saharan Africa, basic infrastructure such as potable water supply and improved sanitation, which are fundamental to the interruption of schistosomiasis transmission, are still lacking (Fenwick et al. 2009). Therefore, exposure to the sources of infection is almost unavoidable and the disease has continued to spread to new geographical areas.

Preventive chemotherapy, involving the large scale administration of praziquantel, is the recommended strategy for schistosomiasis control (Fenwick et al. 2009). This strategy is effective in reducing the intensity of infection (worm load) by 85-95% in most cases, thereby resulting in morbidity control and reduced transmission rates in endemic areas (Chitsulo et al. 2000, Gray et al. 2010). However, since praziquantel is neither able to act on immature schistosomes nor prevent re-infection (Cioli & Pica-Mattoccia 2003), repeated treatment is required at regular intervals. In most cases, the disease prevalence in endemic settings is known to return to pre-intervention levels within 18-24 months of discontinuing treatment (Gray et al. 2010). Despite this, however, preventive chemotherapy has remained the only realistic approach for controlling schistosomiasis, especially in resource-poor settings such as sub-Saharan Africa where the provision of potable water and improved sanitation is still woefully inadequate (Fenwick et al. 2009, WHO 2013*b*).

In a bid to improve the effectiveness of chemotherapy interventions in endemic areas, control programmes have primarily employed an age-targeted strategy of implementation that focuses mainly on children. Therefore, the threshold prevalence levels of high ($\geq$ 50%), moderate ($10\% - 49\%$) or low ($< 10\%$) which are used in establishing modalities for the implementation of chemotherapy are based on rapid assessment of infections in subsets of school children at the district level within specific endemic settings (WHO 2006*b*). This strategy of implementation is in keeping with the World Health Assembly's Resolution 54.19 that recommended the regular treatment of at least 75% of school-aged children in highly endemic schistosomiasis settings by the year 2010 (WHO 2006*a*, Utzinger et al. 2009, WHO 2013*b*).

Therefore, large scale chemotherapy intervention programmes across sub-Saharan Africa, including campaigns by the Schistosomiasis Control Initiative (SCI), have adopted this strategy and made school-aged children (6-15 years) the primary focus of intervention. Treatment is, however, extended to cover entire communities in areas where the baseline threshold prevalence levels are found to be 50 percent or higher (Lammie et al. 2006).

The effectiveness of these chemotherapy intervention programmes, therefore, depends on their ability to achieve high coverage rates in populations that require treatment. However, even though this implementation strategy is able to prioritize the allocation of interventions by focusing mainly on the high risk groups, studies have shown limitations in their effectiveness to attain full coverage. For instance, 50% of the target population in endemic areas is estimated to be missed by intervention programmes that employ this strategy of implementation (Utzinger et al. 2009).

Moreover, the World Health Assembly's Resolution 54.19 of extending regular treatment to 75% of school-aged children in endemic areas by the year 2010 was not realised. This, coupled with the fact that the treated population in sub-Saharan Africa as of 2011 only represented 9.8% of the population requiring treatment, goes to show that the scaling up and proper targeting of chemotherapy interventions still remain major challenges (WHO 2013*a*). Despite these issues with coverage, however, studies in post-intervention areas are also increasingly highlighting the fact that non-infected people tend to constitute the majority of the population that receive chemotherapy during intervention campaigns (Fenwick et al. 2009, Utzinger et al. 2009).

These limitations have, therefore, raised concerns about how the effectiveness of the existing implementation strategy might be improved in order to extend coverage to populations that actually require treatment. For example, mass treatment regimes in some of the countries that were initially targeted for intervention by the Schistosomiasis Control Initiative, such as Burkina Faso and Niger, have not resulted in reduced prevalence levels (Gray et al. 2010). This, therefore, suggests chemotherapy interventions that rely on the current implementation strategy may not be having any significant impact on transmission rates in certain endemic settings.

Indeed, the use of school-aged children as epidemiological indicators of the disease prevalence has been debated in some circles. Basing their argument on field surveys involving preschool-aged children in Ghana (Bosompem et al. 2004) and Uganda (Odogwu et al. 2006), Stothard and Gabrielli (Stothard & Gabrielli 2007) have pointed out that since preschool-aged children (1-5 years) also tend to harbour high infection rates, modalities that are established using school-aged children may end up under-estimating the actual

prevalence at the community level. However, while infection in preschool-aged children is recognised, they are still excluded from preventive chemotherapy interventions mainly as result of challenges related to the licensing limitations of praziquantel (Stothard et al. 2013).

Further evidence in support of the inadequacy of the school-based approach of establishing modalities is provided by studies in Egypt that found infection rates to be higher in non-enrolled school-aged children as compared to those who were enrolled in schools (Husein et al. 1996, Talaat & Evans 2000). Globally, sub-Saharan Africa accounts for half of all non-enrolled school-aged children and as of 2010 had a school drop-out rate of 42%, which is also the highest in the world. This is compounded further by the fact that 32% of enrolled children in that region are above the expected enrolment age of 6 years at the beginning of their compulsory education (UNESCO 2012). This, therefore, emphasises the need for control programmes to look beyond the school-based approach, especially in settings where interventions are aimed towards morbidity control as well as the interruption of transmissions (WHO 2013*b*).

For instance, the impact of chemotherapy interventions is most realised when the initial administration of praziquantel is followed-up by at least two subsequent treatments. In most cases, the prevalence of heavy-intensity infections is known to reduce to $< 5\%$ when coverage of about 75% of the target population is sustained through repeated treatment over a number of years. The strategy for administering these follow-up treatments involves an annual and biennial re-administration in high and moderate risk areas, respectively; and twice during the period of primary schooling in low-risk areas (WHO 2013*b*). However, the high school non-enrolment and drop-out rates in the region suggest these benefits would remain unattainable until the focus of these interventions is extended beyond schools to the community level.

The development an alternative implementation strategy that overcomes the aforementioned limitations has, however, proved elusive. Though the adoption of a blanket mass chemotherapy approach may provide a solution, it has not been considered an option for the following reasons. Praziquantel still remains the most expensive anthelmintic drug, despite the expiration of its patency. Hence, its procurement in endemic countries

has only been made possible through external funding by donor agencies (WHO 2010). Therefore, the judicious use of available drugs would ensure the sustenance of ongoing chemotherapy interventions through follow-up treatments.

Schistosomiasis is, however, characterised by a unique epidemiological feature where the transmission of infections tend to cluster, rather than occur at random within endemic settings. This feature, which is often referred to as the focalisation of risk, may vary at different spatial scales (Brooker 2007). Moreover, different factors are known to influence this focalisation of risk at the different spatial scales (Brooker 2007). At the local level, setting-specific micro-level factors that determine human exposure patterns to the sources of infection may influence the degree and patterns of aggregation of infected cases (Bruun & Aagaard-Hansen 2008). Therefore, using these setting-specific heterogeneities in transmissions as the focus of intervention could provide an option to improving the effectiveness of chemotherapy interventions and overcoming some of the current challenges with coverage.

The present study, therefore, investigates the setting-specific variations in the transmission of schistosome infections and the micro-level factors that influence these heterogeneities, using data from three endemic shoreline communities of the Volta Lake in Ghana. Our primary objectives are: to model the risk of *Schistosoma haematobium* and *S. mansoni* infections in each of our study communities; and to assess the residual spatial effects due to micro-level factors specific to each of the study communities. Therefore, any observed residual spatial variation within any of these communities would signify heterogeneities in risk and potential targets for the administration of chemotherapy interventions.

2 Materials and Methods

This study is, therefore, based on the following hypotheses: that the transmission of schistosome infections within endemic settings tends to assume heterogeneous patterns, rather than occur at random. The degree of this heterogeneity also tends to vary within different endemic settings in response to specific micro-level factors that govern human exposure patterns. Therefore, these heterogeneous patterns of transmissions could serve as effective targets for the implementation of chemotherapy interventions within different endemic settings.

2.1 Study Sites

To investigate these hypotheses, we used pre-collected point-referenced data resulting from community-level cross-sectional surveys on the physically active and permanent residents of three endemic shoreline communities of the Volta Lake in Ghana (Figure 8). Since the original sampling strategy focused on inhabitants in the 5 - 60 years age range, our analysis was also restricted to that age range. Ghana is considered one of the most highly endemic settings for schistosomiasis in sub-Saharan Africa and the vast majority of these infections is concentrated in the shoreline communities of the Volta Lake which covers a total area of 3,275 square miles, stretching across a shoreline of 3,106 miles (WHO 2013*b*, Steinmann et al. 2006, WHO 1987). Though these shoreline communities are predominantly rural, their levels of development also tend to vary. Therefore, to allow an effective investigation of our study hypotheses, our selected data from the pre-existing database were chosen to reflect communities with different levels of development and chemotherapy intervention histories. In the next two subsections, we present a description of our study communities at baseline.

2.1.1 Alabonu

Alabonu was still a pre-intervention site for schistosomiasis as of the time of the data collection. However, most of the school-aged children in the community had received a single 500mg dose of mebendazole as part of a national school-based de-worming programme for

the soil-transmitted helminth (STH) infections. The de-worming programme took place two years prior to our data collection. The community, which had a total population of about 800 people, lacked potable water supply, improved sanitation facilities and good access roads. Therefore, exposure patterns were virtually constant among the physically-active inhabitants. Moreover, the lake served both domestic and commercial purposes.

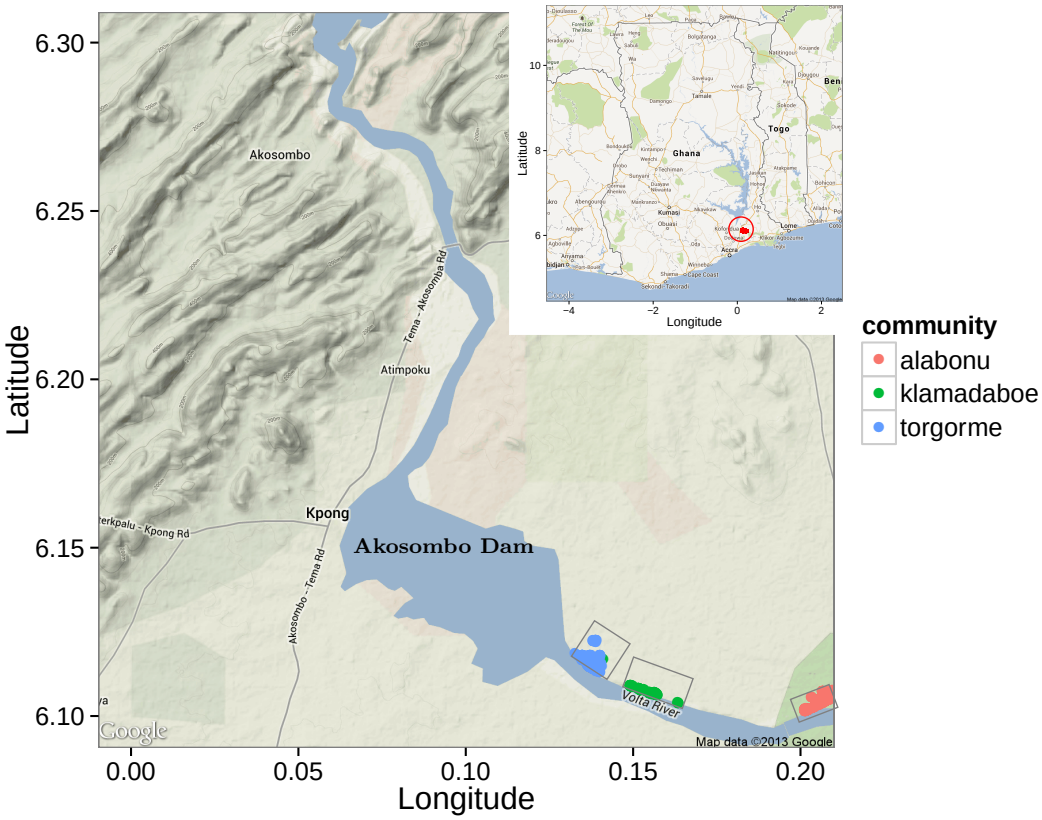


Figure 8: Map of the study sites showing the locations of the three communities relative to the Akosombo Dam and the Volta Lake (Main). Inset: Map of Ghana showing the locations of the study sites (circled). R Packages used in map preparation: ggmap (Kahle & Wickham 2013).

2.1.2 Torgorme and Klamadaboe

The two other communities, Torgorme and Klamadaboe, were contiguous. Torgorme was, however, slightly more developed than our other study sites. Apart from having an access road that granted the inhabitants an alternative means of transportation, Torgorme also had a communal tap, a health centre and public sanitation facilities.

Moreover, unlike our other study sites which were still pre-intervention areas for schistosomiasis, Torgorme was in receipt of sporadic chemotherapy interventions for schisto-

somiasis as part of the Volta River Authority's outreach programmes for selected pilot communities (VRA 2012). The last of such treatments was administered about three years prior to our data collection. Moreover, school-aged children in both Torgorme and Klamadaboe had received treatment for the soil-transmitted helminth infections during a national school-based de-worming programme which occurred two years prior to our data collection.

2.2 Ethical Considerations

The study was approved by the Institutional Review Board of the Noguchi Memorial Institute for Medical Research in Accra, Ghana. Prior to recruiting participants, the chiefs and elders of the chosen communities were asked for approval. Informed consent was obtained from the inhabitants before recruitment into the study. Children were recruited after obtaining their assent and informed consent from their parents or guardians. All infected participants were treated with praziquantel at the end of the study, using treatment guidelines by the World Health Organisation (Montresor et al. 2005).

2.3 Measured Exposures

The measured exposure variables that were considered in this study are presented in Table 2 below. Information on these variables were collected during a questionnaire survey that was conducted at baseline in the study communities. A sample of the questionnaire is attached as Appendix A.

2.4 Study Outcomes

Excreted egg output in urine and stool samples collected during a parasitological survey was used as proxy for schistosome infections. Our study outcomes, *S. haematobium* and *S. mansoni* infections, were therefore defined as: 1) the presence of one or more *S. haematobium* ova in 10 ml of filtered urine samples; or 2) the presence of *S. mansoni* eggs in one gram of stool sample processed by the Kato-Katz technique (WHO 1985). Both infections were detected by microscopy.

Table 2: Exposure variables measured during question survey.

Variable	Description	Factor Levels	Questionnaire Responses
Age	Age of participants		5-60 years
Water source	Main source of water for domestic use		"Pipe borne water", "Pipe"
		Lake	"River", "Volta River"
		Both Tap and Lake	"Pipe borne water and river", "Pipe borne water or river"
Frequency	Frequency of contact with the Lake	1-3 times per week	"1-3 times", "4-6 times"
		1 per month	"Once a month", "Once a year"
		Daily	"Daily"
STH treatment	History of treatment for soil-transmitted helminthiases (STH)	Once	"Once"
		Twice	"Twice", "Thrice"
		Never	"Never", "Not certain"
Last STH treatment	When treatment for STH was last administered	1 year	"Up to 1 year ago"
		2 years	"Up to 2 year ago"
		3 years	"Up to 3 year ago", "More than three years ago"
Source of anthelmintics	Where treatment for STH was administered	Bought	"Drug store"
		Hospital	"Dispensary"
		VRA	"Volta River Authority staff", " [b] Korean NGO", "National de-worming programme"
Awareness	Knowledge of schistosomiasis	Very well	"Yes, very well"
		Not very well	"Yes, but not very well"
		No	"No"
Haematuria	Is blood in urine is a symptom of schistosomiasis?	No idea	"No idea"
		True	"True"
		False	"False"
Dysuria	Is painful urination a symptom of schistosomiasis	No idea	"No idea"
		True	"True"
		False	"False"
Bloody diarrhoea	Is blood in stool a symptom of schistosomiasis	No idea	"No idea"
		True	"True"
		False	"False"

[b]Korean NGO: A Korean Non-Governmental Organisation.

Table 2 continued.

Variable	Description	Factor Levels	Questionnaire Responses
History of haematuria	Ever experienced haematuria?	No Yes	"No" "Yes"
History of dysuria	Ever experienced dysuria?	No Yes	"No" "Yes"
History of bloody diarrhoea	Ever experienced had blood in stool?	No Yes	"No" "Yes"
Management of symptoms	How participant manages symptoms of schistosomiasis	Did nothing Bought medicine Hospital Herbs	"Nothing was done" "Bought medicine" "Went to the hospital" "Used herbal medicine"
Schistosomiasis treatment	History of treatment with praziquantel	Once >twice Never	"Once" "Twice", "Thrice" "Never", "Not certain"
Last Schistosomiasis treatment	When treatment with praziquantel was last administered	1yr >2yrs	"Up to 1 year ago" "Up to 2 year ago", "Up to 3 years ago"
Source of the praziquantel	Where treatment was administered	Hospital VRA Bought School	"Dispensary", "Hospital staff" "Volta River Authority staff" "Bought medicine" "At school"
Water contact activities	Activities that results in exposure to the Lake	Water collection Washing Recreational Economic	"Fetching water" "Washing", "bathing" "Swimming" "Fishing", "canoe boarding", "sand winning", "wading through"

2.5 Exploratory Analysis

2.5.1 Criteria for Inclusion in the Analysis

Participants were considered for inclusion in the analysis if they had submitted samples for at least one of our study outcomes as well as exposure data (Table 3).

Table 3: Total number of participants who were considered for inclusion in the analysis. The inclusion criteria was based on participation in at least one of the parasitology surveys as well as the questionnaire survey.

	Total Number of Participants (%)			
Community	**Both Urine and Stool Samples**	**Urine Samples Alone**	**Stool Samples Alone**	**Total**
Alabonu	512 (59.33)	14 (23.33)	7 (11.48)	533
Klamadaboe	122 (14.14)	5 (8.33)	8 (13.11)	135
Torgorme	229 (26.54)	41 (68.33)	46 (75.40)	316
Total	**863**	**60**	**61**	**984**

2.5.2 Assessing Non-Linearity in Age

Based on a priori knowledge, we expected non-linear relationships between age and our study outcomes, the prevalence of *S. haematobium* and *S. mansoni* infections (Y_i). Studies in different endemic settings have consistently reported peak prevalence rates in children aged 10-14 years whilst adults usually tend to show a decline in prevalence rates. The declining rates in adults is, however, less pronounced for *S. mansoni* infection (Jordan & Webbe 1982). We, therefore, employed the use of non-parametric smoothers, equation 1, in assessing these relationships (Wood 2006). Reasonable approximations about the forms of dependence on age in our formal models were then based on the estimated smoothers.

$$\begin{aligned} \log\left(\frac{\pi_i}{1-\pi_i}\right) &= f\left(\mathrm{X}_i\right) \\ Y_i &\sim \text{Bernoulli}(1, \pi_i) \end{aligned} \tag{1}$$

where π_i denotes the probability that the observed outcome in the i^{th} participant was positive and $1-\pi_i$ denotes otherwise; X_i represents the continuous-valued covariate, age, whilst $f(\mathrm{X}_i)$ is the smoothing function for age in the Generalised Additive Model.

2.6 Confirmatory Analysis

To reiterate our study assumptions, we are investigating the hypothesis that the transmission of schistosome infections may exhibit setting-specific variations. Therefore, given our outcome data on the presence or absence of *Schistosoma haematobium* and *S. mansoni* infections in our study population, we proceeded to investigate the factors that influenced the risk of infection with either outcome separately for each of our study sites. Following our discussions in sections 2.3 and 2.5.2 above, the measured exposures that were considered in this study comprised of categorical covariates as well as the continuous-valued covariate, age, for which every participant had a unique value. Therefore, to determine which of these exposures increased the probability of infection with either outcome in any given participant, we formulated our model as follows.

2.6.1 Model Formulation

The effect of the factors that influenced the observed outcome in the ith participant was modelled by assuming there is as an unknown probability, π_i, of acquiring an infection. Therefore, higher values of this probability, π_i, would signify higher chances of being infected and vice versa (Collett 2003). Moreover, if the infection status of the ith participant were denoted by a random variable, $Y_i \in \{0, 1\}$, then $P(Y_i = 1) = \pi_i$ would signify the probability that an infection is observed whilst $P(Y_i = 0) = 1 - \pi_i$ would denote otherwise. Therefore, Y_i follows the Bernoulli probability distribution which is given by equation 2.

$$P(Y_i = y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i} \tag{2}$$

where

$$y_i \in \{0, 1\}$$

Following the Bernoulli convention, the expected mean, $E(Y_i)$, and variance of Y_i are given by equation 3. The systematic component of the model, which is specified as a function of the measured exposures, is given by equation 4.

$$\begin{aligned} E(Y_i) &= \pi \\ var(Y_i) &= \pi(1 - \pi) \end{aligned} \tag{3}$$

$$\eta_i = \beta_1 + \beta_2 \mathbf{X}_i \tag{4}$$

where eta, η, is the predictor function, β_1 and β_2 are the intercept and vector of unknown regression parameters, respectively whilst $\mathbf{X}_i$ denotes the vector of measured exposures for the ith participant.

To define the relationship between π and η, a function that restricts the values of η between 0 and 1 is required. Of the available functions that serve the purpose, the logit link function is commonly used in linking π and η (Zuur et al. 2013). Hence, the relationship between π and η was defined by equation 5 which limits the fitted values of π within a 0-1 range.

$$logit(\pi_i) = \beta_1 + \beta_2 \mathbf{X}_i \tag{5}$$

$$\pi_i = \frac{e^{\beta_1 + \beta_2 \mathbf{X}_i}}{1 + e^{\beta_1 + \beta_2 \mathbf{X}_i}} \tag{6}$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) \quad = \beta_1 + \quad \beta_2 \mathbf{X}_i \tag{7}$$

2.6.2 Estimation of the Unknown Regression Parameters

With our models specified up to the point of the unknown regression parameters, β, we next proceeded to estimate these parameters. For generalised linear models, the estimation of the unknown regression parameters is effected with the maximum likelihood method. Using the data and the chosen model, the maximum likelihood method estimates values of the regression parameters that make the data most likely (Crawley 2014). Therefore, the process of parameter estimation involves the specification a joint likelihood criterion for the data which is then maximised as a function of the regression parameters (Zuur et al. 2013).

The estimation process begins with the formulation of the likelihood function, based on the probability distribution function. Moreover, in order to simplify the maximisation process, the logarithm of the likelihood function is employed in making the likelihood criterion additive. For the Bernoulli probability distribution, equation 8, the likelihood

and log likelihood functions are given by equations 9 - 10.

$$f(y_i; \beta_1, ...\beta_k) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{8}$$

$$L = \prod_{i=1}^{N} f(y_i; \beta_1, ...\beta_k) = \prod_{i=1}^{N} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{9}$$

$$log(L) = \sum_{i=1}^{N} log\left(\pi_i^{y_i}(1-\pi_i)^{1-y_i}\right) \tag{10}$$

$$= \sum_{i=1}^{N} y_i \times log(\pi_i) + (1-y_i) \times log(1-\pi_i)$$

The estimation of values of the regression parameters that maximise the log-likelihood can be effected with either frequentist or Bayesian estimation techniques. The chosen technique optimises the log-likelihood by obtaining the first order derivatives which are then set to zero. A suitable optimisation routine is then employed in solving the resulting equations in order to estimate the regression parameters (Zuur et al. 2013). We fitted a standard binary logistic regression model which estimates the regression parameters by the frequentist technique.

2.6.2.1 Model Selection

Simple logistic regression analyses were initially performed to assess the worth of each measured covariate. These covariates were then entered into the model in the order of their contribution to the overall deviance, as judged by a deviance-based test. The relative effects of the terms in the multiple logistic regression model were also assessed by a deviance-based test (Hilbe 2009). Therefore, terms that were significant on entry into the model but whose effects became non-significant after adjusting for the effects of other terms were considered for exclusion. The exclusions of any such terms were, however, effected only when their omissions did not significantly increase the overall deviance of the model. The resulting model, after the backwards selection of terms, became our main effects model for the risk of the schistosome infections.

Using selected terms in the main effects model, potential two-way interactions were investigated. The effects of these interaction terms were assessed by computing the deferences in deviance between the nested model, i.e. the main effects model, and

each of the full models containing the interaction terms, in addition to the terms in the main effects model. Statistically non-significant interaction terms were dropped if their exclusions did not significantly increase the overall deviance of the model. The resulting model became our provisional model for the risk of schistosome infections.

2.6.2.2 Goodness-of-Fit Check

2.6.2.2.1 Assessing Over-dispersion

Following recommendations by Zuur et al. 2013 that over-dispersion in Bernoulli GLM be defined by the Pearson dispersion statistic, equation 11 was used in assessing our provisional models for extra-dispersion. Our cut-off limit for over-dispersion was set at 1 (Zuur et al. 2013).

$$\frac{\sum(y_i - \pi_i)}{N - k} \tag{11}$$

where y_i is the observed value of the outcome in the i^{th} participant; π_i is the probability that infection was observed in the i^{th} participant; N denotes the number of observations in the model and k is the number of parameters in the model, including the intercept.

2.6.3 Model Validation

2.6.3.1 Assessing the Homoscedasticity Assumption

In assessing our provisional model for non-constance error variance, the Pearson residuals, equation 12, were sorted into groups of 10 based on the order of their corresponding sorted fitted values (Zuur et al. 2011). The mean of the Pearson residuals in each of these groups was then computed and plotted against the fitted values to assess homoscedasticity. This approach of assessing the homoscedasticity assumption in binary logistic regression models, however, tends to work better for large data sets (Zuur et al. 2011).

$$\epsilon_i = \frac{Y_i - \hat{\pi}_i}{\sqrt{\hat{\pi}_i\,(1 - \hat{\pi}_i)}} \tag{12}$$

where Y_i is the observed response for the ith participant, $\hat{\pi}_i$ is the fitted value corresponding to Y_i and $\sqrt{\hat{\pi}_i\,(1 - \hat{\pi}_i)}$ is an estimate of the standard deviation of the raw residuals.

2.6.3.2 Assessing the Assumption of Independence

In line with our study hypothesis, we expect the transmission of schistosome infections to assume heterogeneous, rather than random patterns within each of our study communities. Such non-random patterns are, however, often the results of complex interactions between the measured covariates and specific micro-level factors that act in conjunction with the common risk factors. Therefore, in assessing the effect of these micro-level factors on the validity of our models, certain assumptions were made. These assumptions are discussed in the next two sections.

2.6.3.2.1 Assumptions Regarding the Effects of the Micro-Level Factors

We began by assuming that the micro-level factors were unobserved latent factors specific to each of our study sites. These factors were also thought to have acted in conjunction with the measured covariates to influence exposure patterns in the study communities. Moreover, in order for the collective effects of these micro-level factors to be evident, we further assumed that they were mostly activities that may have occurred in a socio-temporal context among groups of individuals, rather than as discrete activities that were performed at the individual-levels.

For instance, the infective stages of the schistosome parasites, the cercariae, are known to exhibit diurnal rhythms where their peak densities in the freshwater habitats occur between noon and 15:00 GMT (Farooq & Mallah 1966). Therefore, human exposure activities that occur within this peak period are bound to be associated with higher risk of infectivity. Moreover, since the water bodies that serve as sources of transmission tend to be major sources of livelihood in endemic settings, human exposure patterns may also be linked to the major economic activities specific in these settings.

Therefore, if similar levels of infectivity were assumed across the different points of exposures, then the time of exposure may potentially contribute to the varying risk of infection in the human host population. Hence, we could logically assume that for any given endemic setting, groups of inhabitants involved in the same occupational activities may experience similar risk due to similarities in the time of exposure.

2.6.3.2.2 Assumptions Regarding the Residuals

The residuals of our provisional model represented the variation in the risk of schistosome infections that was not accounted for by the measured covariates (Arlinghaus 1996). Hence, it follows that these residuals represented the effect of the micro-level factors as well as the random noise in the data. Any apparent patterns in these residuals may, therefore, signify the clustering of risk due to the effects of micro-level factors acting in specific parts in any of our study sites. We, therefore, proceeded to investigate heterogeneities in the patterns of schistosome transmissions by conducting an assessment for spatial autocorrelation in the standardised point-referenced residuals.

2.6.3.2.3 Principle of Spatial Autocorrelation

The principle of spatial autocorrelation is based on Tobler's first law of geography which states that "*everything is related to everything else, but near things are more related than far things*" (Waller & Gotway 2004). This principle, therefore, captures the distance decay concept in spatial statistics where nearby objects are thought to share more similar attributes, and hence could no longer be considered as being independent of each other (Arlinghaus 1996).

In the context of this study, the clustering of risk would be expected to be evident in the point-referenced residuals if the inhabitants of any of our study communities, who also happened to live in close proximities to each other, were exposed to similar micro-level factors. Therefore, using the standardised Pearson residuals, which represented the variation in the risk of infections that was unaccounted for by our provisional model, we preceded to conduct an assessment for spatial autocorrelation. In the next section, we focus on the classical geostatistical convention for effecting the decomposition of our provisional model into covariate information and residuals.

2.6.3.2.4 Classical Geostatistical Concepts

Following classical geostatistical convention, residuals (z_i) from the various sampled locations, $\mathbf{s}$, within our study communities could be regarded as samples of a single realisation

of an underlying random and spatially continuous process, **Z**. These observed realisations are therefore used in drawing statistical inferences about the random function, **Z**. The spatial distribution of **Z** is specified by the mean, μ and covariance or variogram i.e. the first two moments (equation 13) (Gelfand et al. 2010, Waller & Gotway 2004).

$$E\left[\mathbf{Z(s)}\right] = \mu \tag{13}$$
$$Cov\Big(\mathbf{Z}(\mathbf{s}_j), \mathbf{Z}(\mathbf{s}_k)\Big) = C(\mathbf{s}_j - \mathbf{s}_k)$$

where $C(.)$, the covariance function, measures the spatial autocorrelation between sampled locations $\mathbf{s}_j$ and $\mathbf{s}_k$. Under the assumption of second-order stationarity for the random function, **Z**, μ is independent of location and the covariance only depends on the separation distance between $\mathbf{s}_j$ and $\mathbf{s}_k$(Waller & Gotway 2004).

2.6.3.2.5 Variogram Analysis

The variogram, defined by equation 14, is the geostatistical tool for measuring the spatial dependence between the residuals, z_i, at sampled locations $\mathbf{s}_j$ and $\mathbf{s}_j$ + h. The choice of appropriate lags, h, is conventionally based on the mean distance between pairs of sampled locations (Figure 9) (Myers 1997). The dependence in the spatial process is evidenced by small spatial lags, h, between sampled locations due to the similarity in the values of residuals, z_i (Figure 10) (Zuur et al. 2007). In computing the empirical variogram, specific functions in the `R geoR` package by Jr & Diggle 2012 were employed. The variogram was computed for 13 spatial lags with a lag tolerance of $\pm$ 22.5°.

$$\gamma(\mathbf{h}) = \frac{1}{2p(\mathbf{h})} \sum_{\alpha=1}^{p(\mathbf{h})} \{z_i(\mathbf{s}_j) - z_j(\mathbf{s}_k)\}^2 \tag{14}$$

where $p(\mathbf{h})$ denotes the numbers of pairs of Pearson residuals that are separated by **h**, the spatial lag, **h**, is the distance separating any given set of pairs of residuals while z_i and z_j are the residuals for the ith and jth participants at locations $\mathbf{s}_j$ and $\mathbf{s}_k$, respectively.

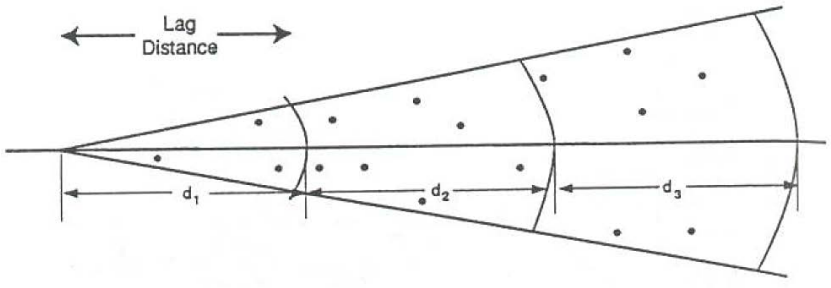


Figure 9: A schematic representation of the lag distances, d_1 - d_3, and residuals that occur within each lag (adapted from Myers 1997 with permission of the rights holder, John Wiley and Sons).

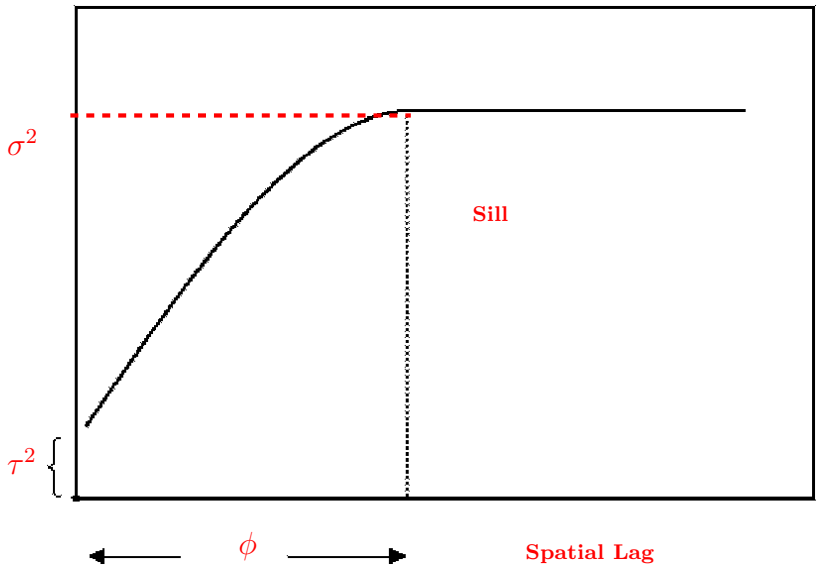


Figure 10: An example of a typical variogram (adapted from Waller & Gotway 2004 with permission of the rights holder, John Wiley and Sons). The variance of the spatial process, Y, is given by the sum of the measurement error variance, τ^2, and the signal variance, σ^2 whilst ϕ denotes the range within which residuals are spatial correlated.

2.6.3.2.6 Directional Dependence

As discussed in section 2.1 above, the pre-intervention sites also lacked any of the operational components of transmission control. Hence the patterns of schistosome transmission in those communities could be attributed primarily to the distribution and abundance of the intermediate aquatic molluscan hosts (Brooker 2007). Therefore, if the degree of infectivity were assumed to be constant across all the points of exposure, i.e. the water contact sites, then we could also ideally speculate that the strength of the spatial correlation would be the same in all directions assuming transmissions were autochthonous. Hence, the omni-directional empirical variogram could ideally be regarded as the mean variogram for all spatial directions in such instances.

However, if the degree of infectivity did vary across the different points of exposure, due to site-specific factors that influenced the ecology of the intermediate snail host species, then we could adopt the logic by Vounatsou et al. 2009 in arguing that the spatial dependence would be stronger in the direction of the higher transmission points. This logic would, however, only hold if the different points of exposure were constantly accessed by the same group of individuals and there was limited interaction in exposure patterns across the different sites (this is investigated in Part III of this thesis).

Though the latter argument seems more plausible, it also suggests a potential violation of the isotropy assumption which renders the omni-directional empirical variogram inadequate. We, therefore, computed directional variograms to verify the isotropy assumption and correct for anisotropy where necessary.

2.6.3.2.7 Statistical Significance of the Spatial Dependency

The position of the isotropic empirical variogram within an envelope of random permutations, computed by the random allocation of the residuals to different sampled locations within the study region, was initially used in assessing if any observed spatial trend in the residuals may have occurred by chance (Diggle & Ribeiro 2007). A formal test for trend by Eagle & Diggle 2012, which is based on a null hypothesis of the absence of spatial autocorrelation, was then employed in computing the test statistic and p-value.

Therefore, a statistically significant spatial autocorrelation was interpreted as signifying residual spatial variation due to the effect of the unobserved micro-level factors. Since this had the effect of overestimating the models' precision, the next step in the analysis was to describe the spatial dependency with the appropriate correlation structure and incorporate the correlation into the model. We, however, defer any discussion on the model adaptation to Part II of this thesis.

2.6.4 Alternative Approach

Instead of modelling our binary outcome variable as a Bernoulli distribution, we could alternatively have modelled Y_i as a binomial distribution (equation 15). The spatio-temporal distribution of the probability of schistosome infections could then have been estimated by employing an approach that utilises the variogram in the model such as indicator kriging or a spatial generalised linear model (Cressie 1993, Diggle & Ribeiro 2007).

$$Y_i \sim B\left(n_i, \pi_i\right) \tag{15}$$

where Y_i is the number of infected participants per household and n_i is the number of sampled participants per household, assuming the probability of testing positive for schistosome infection, π_i, is the same for all n_i.

3 Results

Building on our study hypothesis that the transmission of schistosome infections would assume heterogeneous patterns in response to setting-specific micro-level factors, this section examines some of the micro-level factors specific to each of our study sites and presents results of the formal analysis.

3.1 Alabonu

3.1.1 Baseline Characteristics of Study Population

One hundred and sixty-seven people, constituting 32% of the study population, were infected with *S. haematobium*, 60 (12%) with *S. mansoni* and 24 (5%) with both *S. haematobium* and *S. mansoni*. Although school-aged children accounted for most of the *S. haematobium* infections in the community, *S. mansoni* infection was equally distributed among the school-aged population and adults. However, all the observed infections in the pre-school-aged children were *S. haematobium* infections (Table 4).

The community was a pre-intervention site for schistosomiasis. Therefore, any treated schistosomiasis cases in the community were people who had sought treatment on their own. However, 48% of the participants had received or sought treatment for the soil-transmitted helminth infections, and children in the school-aged population constituted 165 (30%) of this treated group (Table 4).

3.1.1.1 Micro-Level Determinants of Infectivity

3.1.1.1.1 Socio-Temporal Context of Exposure Patterns

The infective aquatic larval stages of the schistosome parasites are known to exhibit diurnal rhythms in response to fluctuations in daily temperature, light intensity and velocity of their aquatic habitats. Consequently, the highest density of the infective larval stages, the cercariae, tends to occur in the water between 12:00 - 15:00 GMT (Figure 11) (Farooq & Mallah 1966).

Table 4: Baseline Characteristics of Study Population in Alabonu

Measured Predictor Variables	*S. haematobium* Infection		Total	*S. mansoni* Infection		Total	Coinfection		Total
	Infected (%)	Non-Infected (%)		Infected (%)	Non-Infected (%)		Infected (%)	Non-Infected (%)	
	167 (31.75)	**359 (68.25)**	**526**	**60 (11.56)**	**459 (88.44)**	**519**	**24 (4.68)**	**488 (95.31)**	**512**
Age (years)									
Preschool-aged (5)	9 (1.71)	20 (3.80)	29	0 (0.00)	29 (5.59)	29	0 (0.00)	29 (5.66)	29
School-aged (6 - 15)	121 (23.00)	127 (24.14)	248	31 (5.97)	217 (41.81)	248	18 (3.52)	228 (44.53)	246
Adults (16 - 60)	37 (7.03)	212 (40.30)	249	29 (5.59)	213 (41.04)	242	6 (1.17)	231 (45.12)	237
Sex									
Female	93 (17.68)	240 (45.63)	333	39 (7.51)	288 (55.49)	327	14 (2.73)	310 (60.55)	324
Male	74 (14.07)	119 (22.62)	193	21 (4.05)	171 (32.95)	192	10 (1.95)	178 (34.77)	188
Frequency of water contact									
Daily	158 (30.15)	312 (59.54)	470	58 (11.22)	403 (77.95)	461	24 (4.71)	432 (84.71)	456
1-3 times per week	9 (1.72)	45 (8.59)	54	2 (0.39)	54 (10.44)	56	0 (0.00)	54 (10.59)	54
Source of water for domestic use									
Lake	167 (31.75)	359 (68.25)	526	60 (11.56)	459 (88.44)	519	24 (4.69)	488 (95.31)	512
Other	0 (0.00)	0 (0.00)	0	0 (0.00)	0 (0.00)	0	0 (0.00)	0 (0.00)	0
Water contact points									
Others	10 (1.90)	59 (11.22)	69	3 (0.58)	65 (12.52)	68	0 (0.00)	67 (13.09)	67
Tutukope torkor	41 (7.79)	135 (25.67)	176	29 (5.59)	145 (27.94)	174	11 (2.15)	158 (30.86)	169
Akpeakpe torkor	116 (22.05)	165 (31.37)	281	28 (5.39)	249 (47.98)	277	13 (2.54)	263 (51.37)	276
Schistosomiasis treatment									
Never	164 (31.30)	338 (64.50)	502	60 (11.61)	434 (83.95)	494	24 (4.71)	464 (90.98)	488
Ever	3 (0.57)	19 (3.63)	22	0 (0.00)	23 (4.45)	23	0 (0.00)	22 (4.31)	22
Schistosomiasis treatment in the past year									
No	166 (31.68)	352 (67.18)	518	60 (11.61)	451 (87.23)	511	24 (4.71)	480 (94.12)	504
Yes	1 (0.19)	5 (0.95)	6	0 (0.00)	6 (1.16)	6	0 (0.00)	6 (1.18)	6
STH treatment									
Never	64 (12.21)	205 (39.12)	269	36 (6.96)	230 (44.49)	266	11 (2.16)	250 (49.02)	261
Ever	103 (19.66)	152 (29.01)	255	24 (4.64)	227 (43.91)	251	13 (2.55)	236 (46.27)	249
STH treatment in the past year									
No	73 (13.93)	242 (46.18)	315	42 (8.12)	267 (51.64)	309	14 (2.75)	290 (56.86)	304
Yes	94 (17.94)	115 (21.95)	209	18 (3.48)	190 (36.75)	208	10 (1.96)	196 (38.43)	206

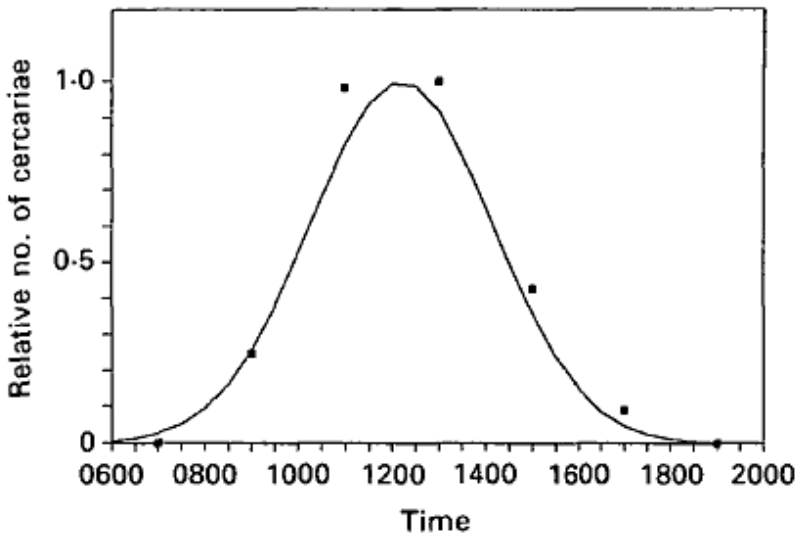


Figure 11: Diurnal variation in the relative density of cercariae discharged by the intermediate snail host of *S. haematobium* (adapted from Chandiwana et al. 1991 with permission of the rights holder, Cambridge University Press.)

Therefore, every exposure event may not automatically lead to an infection (Bruun & Aagaard-Hansen 2008). But rather, the time and degree of exposure may play more significant roles in determining infectivity. Since our provisional models were not adjusted for the time of exposure, there may be residual spatial effects due to interactions between the measured covariates and some of these micro-level factors that influenced infectivity in the community.

For instance, the main economic activities in the community were subsistence farming and fishing. Most of the young adult males also worked on a rice plantation, the Kpong Irrigation Project, that was situated across the Lake. During our field surveys in the community, certain patterns in the time of exposure were observed for different groups of individuals. The next section, therefore, examines some of these variation in water contact patterns from a more epidemiological and socio-economic perspective.

3.1.1.1.2 "Who Goes Where At What Time?"

The question of "who goes where at what time" as posed by (Bruun & Aagaard-Hansen 2008) could best be answered by regarding water contact patterns as processes that occur within the broader context of a series of setting-specific socio-economic activities.

Bearing in mind that there were no realistic alternatives to relying on the Lake for domestic and economic purposes, the rest of this section would focus on the socio-temporal patterns of exposure in the community.

Figure 12: Left: Children taking their bath in the Lake before school. Right: People getting off the canoe in the late afternoon. A woman is seen returning from the cassava mill (circled).

3.1.1.1.3 Common Activities that Influenced Time of Exposure

A normal week day in the community began with the farmers, both males and females, leaving for their farms around 05:00 GMT. The rice plantation farmers took their bath in the Lake around 05:30 GMT and boarded the canoe around 06:30 GMT to leave for work. Around the same time, school children in the community took their bath in Lake before leaving for school (Figure 12). While some of these children were enrolled in the local school, others attended schools in communities across the Lake and had to travel in canoes.

The farmers mainly returned from their farms between 09:30 - 10:00 GMT. While the male farmers normally took their bath almost immediately after getting back home, the female farmers only went to the Lake at that time to collect water. The collected water was then taken home and used in washing some of their farm produce, such as peeled cassava; and for washing clothes (Figure 13).

The rice plantain workers and the students returned to the community around 13:30 and 14:00 - 15:00 GMT, respectively. The late afternoons were usually marked by children swimming in the Lake and women travelling across the Lake to mill their cassava, *Mani-*

Figure 13: Left: Female farmers peeling cassava, while listening our field personnel, around 11:00 GMT. Right: Children returning from school around 14:30 GMT.

hot esculenta, which served as the main ingredient of the staple diet, *akple*. The most notable event linking all these activities is that they all involved some amount of wading in the littoral zone of the Lake at different time points.

Significant departures from these normal routines in the time of exposure occurred during school holidays when the Lake mainly served as a source of recreation for children in the community; and during the rainy season where all activities that involved contact with the Lake were considerably reduced due to the increased risk of drowning. During this period, the inhabitants relied on the rain water for domestic purposes.

3.1.2 *S. haematobium*

3.1.3 Exploratory Analysis

3.1.3.1 Assessing Non-Linearity in Age

An initial assessment of the relationship between age and the prevalence of *S. haematobium* infection suggested that infections were more common among the younger participants ($\leq$ 23 years) whilst the older participants mainly constituted the non-infected population in the community (Figure 14). However, even though Figure 14 suggested a non-linear effect in age, the exact nature of the relationship was not discernible due to the discrete nature of the outcome. Therefore, in line with our discussion in section 2.5.2, we employed the use of non-parametric smoothers by Wood 2006 in estimating a curve that reasonably captured the non-linearity of age on the predictor scale (Figure 15).

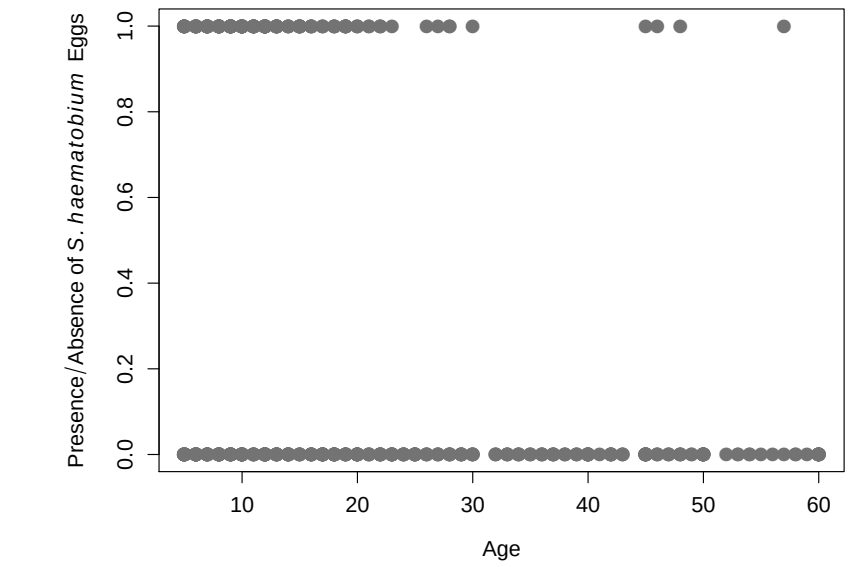


Figure 14: Scatterplot of infection status, as judged by the presence of *S. haematobium* eggs in urine samples, versus age of study participants in Alabonu. The graph mainly suggests *S. haematobium* infections were common in participants aged $\leq$ 23 whilst the older-aged participants mainly constituted the non-infected population.

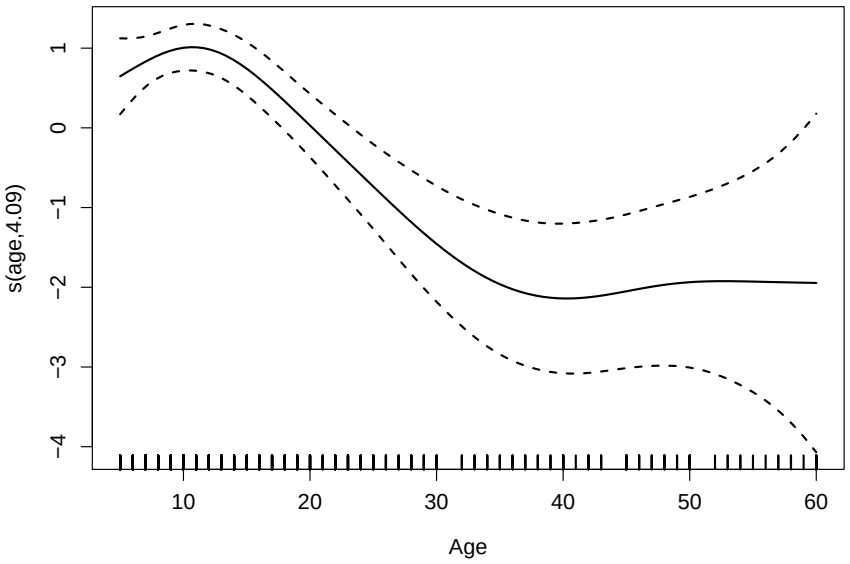


Figure 15: Estimated non-parametric smoother (solid line) showing a non-linear trend in age for the probability of *S. haematobium* infection in Alabonu. The outer dashed lines represent the 95% confidence limits whilst the value, 4.09, in the y-axis label represents the optimum amount of smoothing as determined by the cross-validation process.

3.1.3.1.1 Estimated Non-Parametric Smoother for Age

In line with our earlier observation in section 3.1.3.1, the smoothing function for age, equation 1, estimated a non-linear effect with 4.09 effective degrees of freedom, as opposed to the 1 degree of freedom that would have been produced if the effect were linear (Figure 16 and Table 5). The smoother was significant at the 5% level ($p<0.001$) and accounted for 14 percent of the variation in *S. haematobium* infection. Judging by the shape of the smoother, children aged around 12 years had the highest odds, 0.5, of being infected. Thereafter, a steady decline in the odds of infection occurred until 40 years. The wide confidence intervals above 30 years correspond with the limited number of infections in the older population (Figure 16).

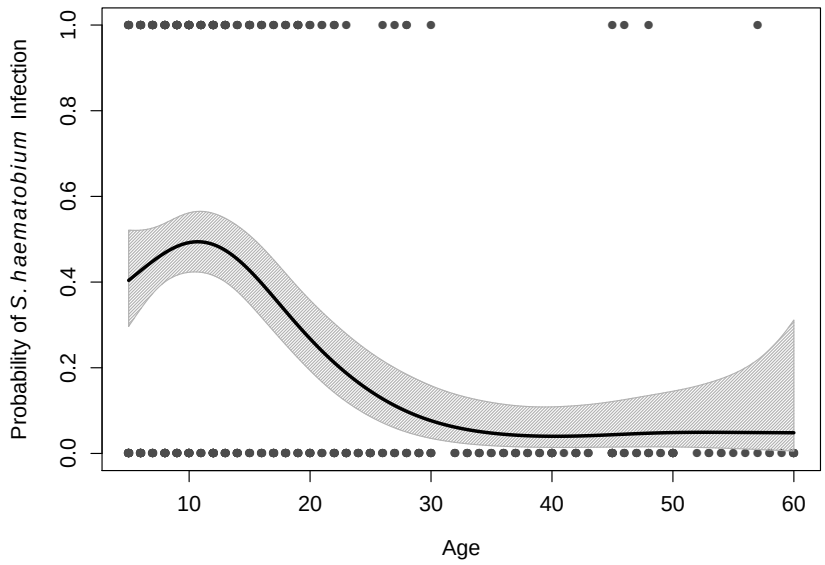


Figure 16: Estimated non-parametric smoother (solid line) for age, equation 1, plotted on the response scale. The grey dots represent the observed prevalence of *S. haematobium* infection in Alabonu.

3.1.3.1.2 Modelling the Effect of Age

If the cross-validation process that estimates the optimum amount of smoothing had produced 1 effective degree of freedom, we could have proceeded to fit age as a linear term in our model (equation 15) without risking any misspecification problems. However, since the evidence we have so far established points to a non-linear effect in age, we were confronted with the task of either fitting a semi-parametric model (equation 16) or programming our own spline that could be incorporated directly into the already specified generalised linear model (equation 15). We settled for the latter option.

Table 5: Model summary for the smoothing function of age in equation 1. The *gam* function which was used in estimating the smoother used the Unbiased Risk Estimator (UBRE) score in the cross-validation process to determine the optimum amount of smoothing (Wood 2006). The amount of smoothing is expressed as effective degrees of freedom (EDF) (Zuur et al. 2011).

Approximate Significance of Smooth Terms					Fit Details				
Smooth Term	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance expl.**	**Scale est.**	**N**
s (Age)	4.087	4.993	54.99	0.000	0.145	0.091423	14.20%	1	526

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) \quad = \beta_1 \quad + \quad f\left(\text{Age}_i\right) \quad + \quad \beta_2 \mathbf{X}_i \tag{16}$$

where equation 16 is a generalised additive model in which π_i, β_1, β_2 and $\mathbf{X}_i$ have the same meanings as for equation 15 and $f\left(\text{Age}_i\right)$ is the smoothing function for age.

3.1.3.1.3 Principle Behind the Smoothing Spline

Conventionally, the smoothing spline is obtained by dividing the X-gradient into a given number of segments and fitting a model to each of the segments. The fitted values per segment are then joined together to form the smoother (Wood 2006, Zuur 2012). Therefore, if we were to revisit equation 1 where the smoothing function for age is given by $f(\text{X}_i)$, then $f(\text{X}_i)$ could be regarded as consisting of basic units, $b_j\left(\text{X}_i\right)$, such that:

$$f\left(\text{X}_i\right) = \sum_{j=1}^{p} \beta_j \times b_j\left(\text{X}_i\right) \tag{17}$$

where p is the total number of segments that make up $f(\text{X}_i)$ and β_j is the regression coefficient obtained by fitting the j^{th} segment in the model.

3.1.3.1.4 Manual Programming of the Smoothing Spline

Therefore, to programme our own smoothing spline, we first needed to determine the number of segments, p, that made up $f(\text{X}_i)$. Judging by the shape of the estimated smoother, section 3.1.3.1.1, changes in the effect of age were observed at 12 and 40 years (Figure 15). Hence, those ages could ideally serve as our break-points on the X-gradient. Therefore, by regarding $f(\text{X}_i)$ as a stick that could broken into p number of segments, the broken stick model approach by White et al. 2014 was adopted in defining the segments, j, of $f(\text{X}_i)$ as equation 18. These segments were then fitted individually in our formal model (equation 15).

$$\textbf{age1} = \begin{cases} \text{age}, & \text{if} \quad \text{age} \leq 12 \\ 12, & \text{if} \quad \text{age} > 12 \end{cases} \qquad \textbf{age2} = \begin{cases} 0, & \text{if} \quad \text{age} \leq 12 \\ 12, & \text{if} \quad \text{age} > 12 \quad \text{and} \quad \leq 40 \\ 28, & \text{if} \quad \text{age} > 40 \end{cases} \tag{18}$$

3.1.4 Confirmatory Analysis

3.1.4.1 Modelling the Risk of *S. haematobium* Infection

3.1.4.1.1 Model Selection

The individual effects of all the measured exposures were significantly associated with the risk of *S. haematobium* infections in the community (Table 7). The relative effects of these covariates in the multiple logistic regression model are presented in Table 8. The effects of age2 (OR: 0.89, 95% CI: 0.85, 0.92), water contact site: Tutukope (OR: 2.35, 95% CI: 1.09, 5.43) and Akpeakpe (OR: 4.75, 95% CI: (2.33, 10.54)); and individual history of schistosomiasis treatment (OR: 0.28, 95% CI: 0.06, 0.93) were significantly associated with the risk of *S. haematobium* infection in the community.

Using the deviance-based test as the basis for their exclusions, the following non-significant terms were dropped from the model: sex, frequency of water contact, history of soil-transmitted helminths treatment and age1 (Table 8). Our main effects model for *S. haematobium* infection was, therefore, adjusted for age2, water contact site and history of schistosomiasis treatment.

None of the two-way interactions between the terms in the main effects were statistically significant (Table 6). The deviance-based test further justified the exclusions of these terms from the model. The provisional model was, therefore, given by Table 9.

Table 6: Deviance analysis for interaction between terms in the main effects model for *S. haematobium* Infection in Alabonu. The difference between the null and residual deviances, D_0 - D_1, represents the contribution of each interaction term to the overall deviance of the model.

	Deviance Analysis				
Interaction Terms	***df***	**D_0 - D_1**	**Residual Deviance**	**Residual *df***	**p-value**
Main Effects Model (Null Model)			**538.46**	**518**	
age2 : water contact point	2	1.367	537.09	516	0.505
age2 : schistosomiasis treatment	1	0.981	537.48	517	0.322
water contact point: schistosomiasis treatment	2	2.622	535.83	516	0.270

Table 7: Simple logistic regression models for *S. haematobium* infection in Alabonu. The difference between the null and residual deviance is given by D_0 - D_1.

	Simple Regression Models			Deviance Analysis		
Model Terms	**OR* (95% CI)**	**SE**	**p-value**	**D_0 - D_1**	**df**	**p-value**
Age						
age2	0.88 (0.85, 0.91)	0.017	0.000	88.950	1	0.000
Water contact point	1					
Others	1.79 (0.87, 4.00)	0.386	0.130			
Tutukope	4.15 (2.12, 8.92)	0.363	0.000	28.320	2	0.000
Akpeakpe						
STH treatment in the past year						
No	1					
Yes	2.71 (1.86, 3.96)	0.193	0.000	27.240	1	0.000
Age						
age1	0.87 (0.81, 0.92)	0.034	0.000	18.190	1	0.000
STH treatment						
Never						
Ever	2.17 (1.49, 3.17)	0.192	0.000	16.720	1	0.000
Frequency of water contact						
Daily	1					
1-3 times per week	0.39 (0.18, 0.79)	0.378	0.014	7.110	1	0.008
Sex						
Females	1					
Males	1.60 (1.10, 2.34)	0.192	0.014	6.040	1	0.014
Treatment for schistosomiasis						
Never	1					
Ever	0.33 (0.08, 0.97)	0.629	0.074	4.070	1	0.000
Schistosomiasis treatment in the past year						
No						
Yes	0.42 (0.02, 2.66)	1.099	0.435	0.730	1	0.393

OR* : Odds ratios computed from the exponentiated coefficients of the models.

Table 8: Multiple logistic regression model for *S. haematobium* infection in Alabonu. The difference between the null and residual deviance is given by D_0 - D_1. The p-values indicate the effect each term would have on the overall deviance of the model if terms were excluded sequentially from the model.

	Multiple Regression Model			Deviance Analysis				
Model Terms	**OR (95% CI)**	**SE**	**p-value**	***df***	**D_0 - D_1**	**Residual Deviance**	**Residual *df***	**p-value**
Age								
age2	0.89 (0.85, 0.92)	0.022	0.000	1	88.190	567.74	522	0.000
Water contact point								
Others	1							
Tutukope	2.35 (1.09, 5.43)	0.406	0.035					
Akpeakpe	4.75 (2.33, 10.54)	0.381	0.000	2	24.120	543.62	520	0.000
STH treatment in the past year								
No	1							
Yes	1.56 (0.69, 3.72)	0.426	0.296	1	1.187	542.43	519	0.276
Age								
age1	1.04 (0.96, 1.14)	0.044	0.349	1	1.070	541.36	518	0.301
STH treatment								
Never	1							
Ever	0.82 (0.35, 1.84)	0.422	0.641	1	0.297	541.07	517	0.586
Frequency of water contact								
Daily	1							
1-3 times per week	1.14 (0.46, 2.64)	0.439	0.766	1	0.001	541.07	516	0.982
Sex								
Females	1							
Males	1.32 (0.86, 2.02)	0.217	0.198	1	1.616	539.45	515	0.204
Treatment for schistosomiasis								
Never	1							
Ever	0.28 (0.06, 0.93)	0.676	0.058	1	4.333	535.12	514	0.037

Table 9: Provisional model for *S. haematobium* infections in Alabonu.

Model Terms	OR (95% CI)	SE	p-value
Age			
age2	0.89 (0.85, 0.91)	0.017	0.000
Water contact point			
Others	1		
Tutukope	2.27 (1.06, 5.24)	0.403	0.041
Akpeakpe	4.69 (2.31, 10.35)	0.379	0.000
Treatment for schistosomiasis			
Never	1		
Ever	0.29 (0.07, 0.97)	0.668	0.067

3.1.4.2 Goodness-of-Fit Check

The dispersion statistic, computed as the ratio of the residual deviance of our provisional model, $D_1 = 539.60$, to its residual degrees of freedom, $df_1 = 519$, was 1.04. Since 1.04 is approximately equal to our limit for over-dispersion as discussed in section 2.6.2.2.1, we concluded that the model was not over-dispersed.

3.1.4.3 Model Validation

3.1.4.3.1 Assessing the Homoscedasticity Assumption

The spread of the residuals across the range of their fitted values suggests no obvious violation of the homoscedasticity assumption (Figure 17).

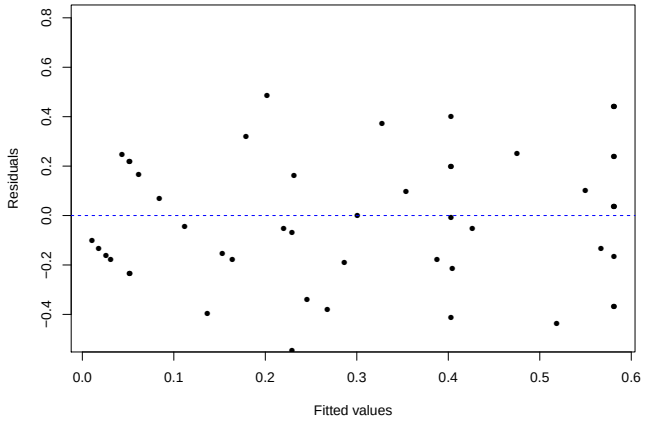


Figure 17: Grouped Pearson residuals, as described in section 2.6.3.1, versus fitted values of the provisional model for *S. haematobium* infection in Alabonu.

3.1.4.3.2 Assessing the Assumption of Independence

A point map of the spatially-referenced residuals, Figure 19, was initially examined for apparent clusters of identical residuals that may potentially violate the assumption of independence. In the context of this study, any such clustering of risk may be indicative of co-variation between the residuals and the un-measured micro-level factors under consideration.

The map, Figure 19, seems to suggest clusterings signifying areas with increased or decreased likelihoods of risk in specific parts of the study community (circled parts of Figure 19). These spatial trends correspond to areas where our provisional model over or under-estimated the risk of infection (Figure 20). However, our interpretation of any such latent trends in the residuals may also have been biased by the cluttering of points resulting from having multiple observations from the same sampled locations. Therefore, some of the points end up being drawn directly over others. As a way if fixing this bias associated with the interpretation of Figure 19, an empirical variogram was employed in quantifying the spatial dependency in the residuals.

3.1.4.3.3 Variogram Analysis

The shape of the omni-directional empirical variogram, Figure 18, mainly shows a flat trend with no sill up to a spatial lag of 0.005° after which it begins to decrease. Hence, it would seem at this stage there is no spatial dependency in the process.

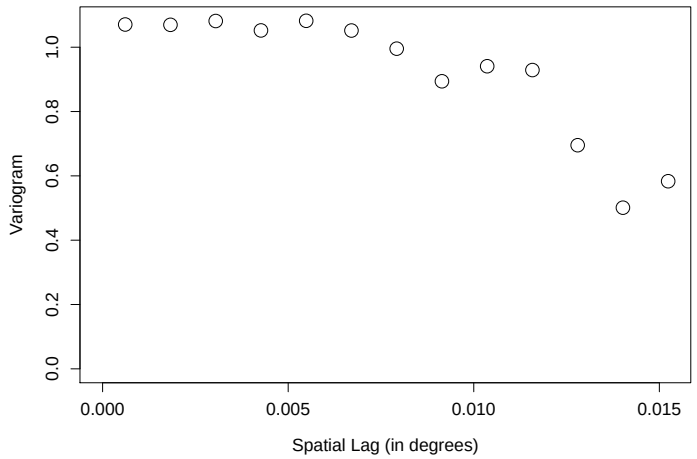


Figure 18: Omni-directional empirical variogram for the Pearson residuals of the provisional model for *S. haematobium* infection in Alabonu.

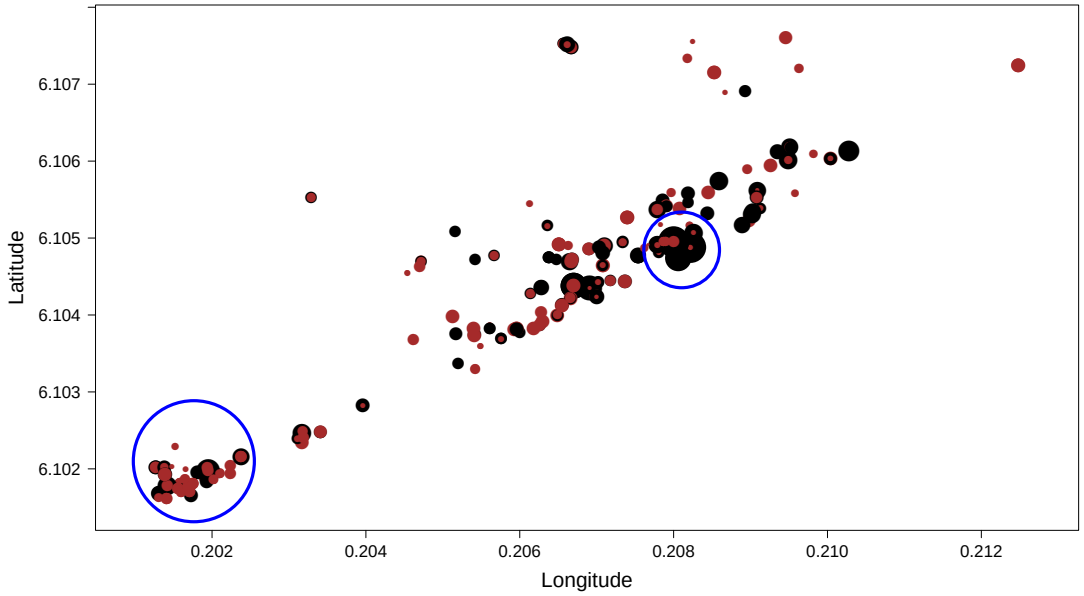


Figure 19: Spatially-referenced residuals of the provisional model for *S. haematobium* infection in Alabonu. The radii of the points are proportional to the absolute values of the Pearson residuals. The red and black points represent the negative and positive residuals, respectively.

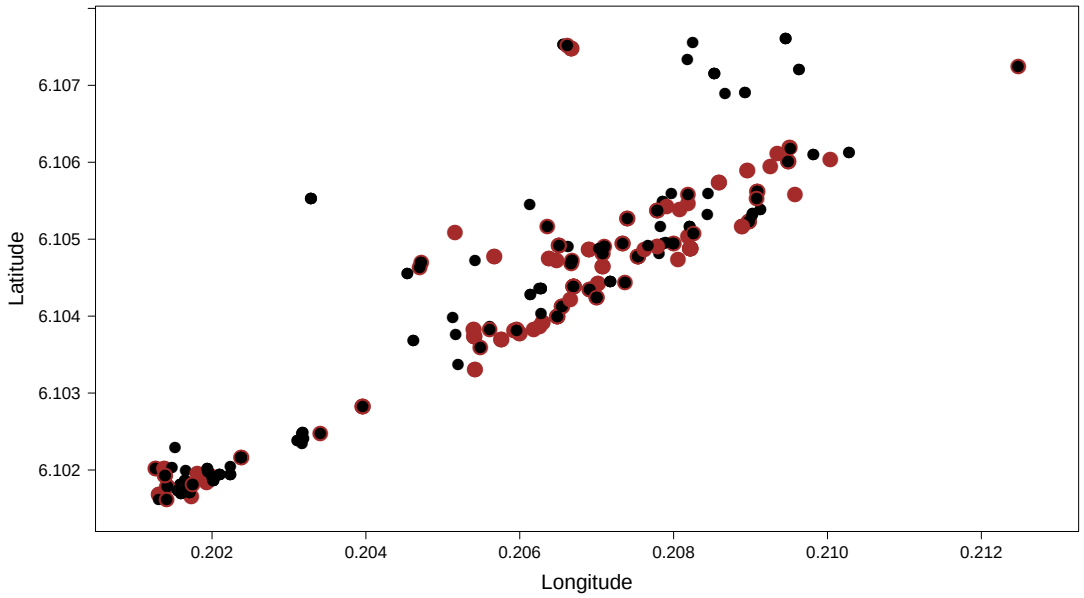


Figure 20: Distribution of *S. haematobium* infection in Alabonu. Each point corresponds to a sampled location i.e location of the household of the sampled participant. The red and black points represent infected and non-infected cases, respectively.

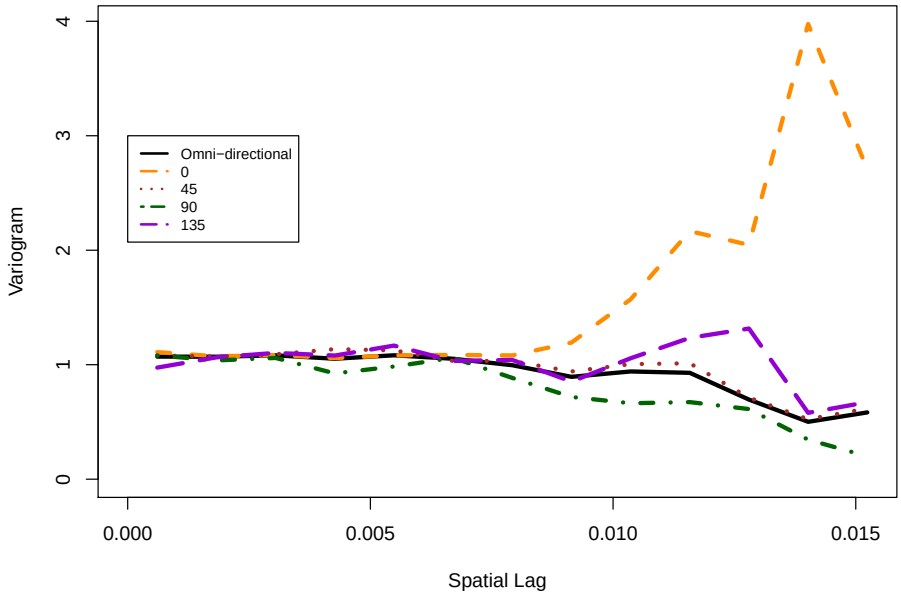


Figure 21: Directional empirical variograms for the 0° , 45°, 90° and 135° spatial directions (with a tolerance angle of ± 22.5° around each of these directions).

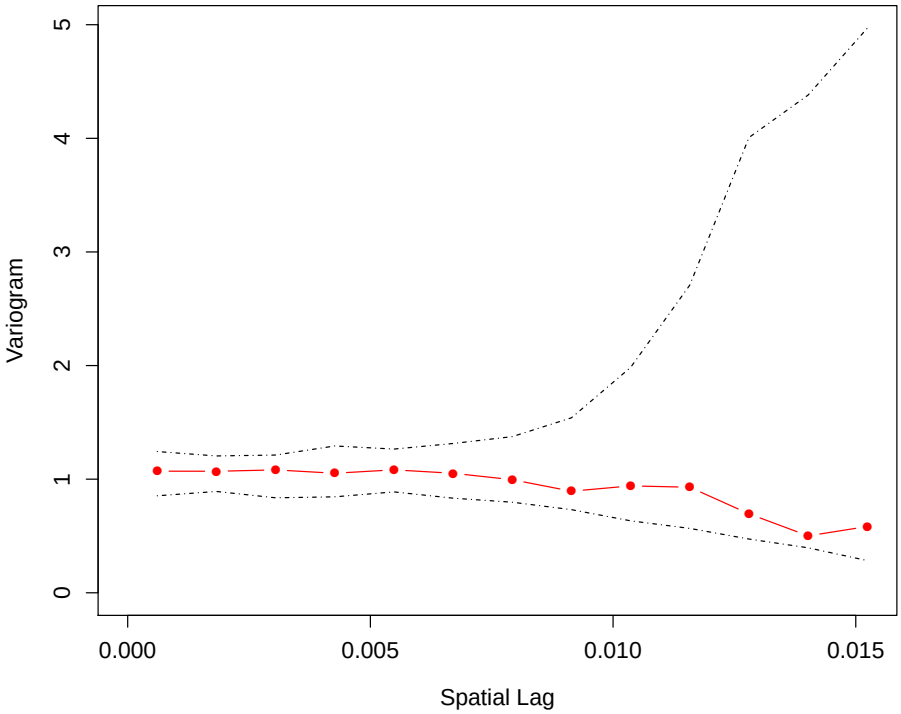


Figure 22: Empirical variogram of the Pearson residuals of the provisional model for *S. haematobium* infections in Alabonu (dotted red line) and the 95% confidence envelope (outer dashed lines). The variogram suggests no evidence of significant spatial autocorrelation in the residuals.

However, we are investigating the hypothesis that the patterns of transmission are heterogeneous in space. Since such heterogeneities may be influenced by the time of exposure as well as the exact point where the different groups of inhabitants frequently experienced their exposures, it was reasonable to assume that any observed spatial trend might also vary with direction. We, therefore, proceeded to verify the isotropy assumption.

The shapes of the computed directional variograms, Figure 21, are mostly flat till a range of around 0.008°, thereby suggesting the absence of any spatial dependence in these directions. The 0°, and to some extent the 135° directions, however, begin to rise and assume different sills after the range 0.008° which may point to different amount of variation in those directions.

However, since none of these directional variograms assumed any of the general shapes that depict the presence of spatial dependency, we chose to use the omni-directional variogram for the rest of the analysis. Both the envelope of random permutations, Figure 22, and our formal tests for trend did not detect any statistically significant spatial dependence, p=0.492.

3.1.4.3.4 Final model for *S. haematobium* Infection

Based on these results, we ruled out the relevance of the effects of any unobserved micro-level factors on the community-level variation in the risk of *S. haematobium* infection, and hence the absence of any significant heterogeneities in the likelihood of infection. Our final model indicates *S. haematobium* infection was explained by age2, the point of frequent exposure and history of schistosomiasis treatment (equation 19).

$$\begin{aligned} \text{logit}(\pi_i) \quad = \quad & 0.29 \quad + \quad 0.89 \quad (\textbf{age2}) \quad + \\ & 2.27 \quad (\textbf{Tutukope}) \quad + \quad 4.69 \quad (\textbf{Akpeakpe}) \quad + \\ & 0.29 \quad (\textbf{treatment}) \end{aligned} \tag{19}$$

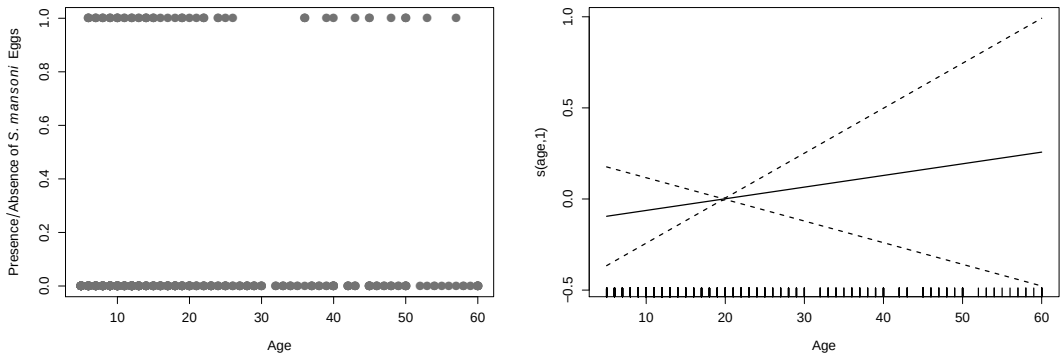
3.1.5 *S. mansoni* Infections

Empirical evidence suggests that infection with any one of the human schistosome species may not necessarily provide any form of immunity against the other schistosomes (Mott et al. 1990). Therefore, even though similar factors may influence the systematic variation in the risk of *S. haematobium* and *S. mansoni* infections, different factors may underlie their transmission dynamics at the micro-level. Hence, we next investigate the factors that influenced the community-level variation in the risk of *S. mansoni* infection in Alabonu.

3.1.6 Exploratory Analysis

3.1.6.1 Assessing Non-Linearity in Age

A scatterplot of the relationship between age and the prevalence of *S. mansoni* infection, Figure 23(a), indicated that though infections were observed across the entire age range for the study, participants aged ≤ 30 harboured most of these infections. To discern the exact nature of this relationship on the predictor scale, the `gam` function in the `R` `mgcv` package by Wood 2006 was employed in estimating a non-parametric smoother that adequately captured the effect in age (Table 10).



(a) Scatterplot (b) Smoother

Figure 23: (a) Scatterplot of the prevalence of *S. mansoni* infection in Alabonu versus age of the study participants. (b) Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the 95% confidence limits (outer dashed lines). The intersection of the three lines is the result of a constraint in the Generalised Additive Models' algorithm that centres smoothers around 0 (Zuur 2012).

Table 10: Summary of model 1 for *S. mansoni* infection in Alabonu. The optimum amount of smoothing was determined by default using the Unbiased Risk Estimator (UBRE) score.

Approximate Significance of Smooth Term					Fit Details				
Smooth Term	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance expl.**	**Scale est.**	**N**
s (Age)	1	1	0.49	0.484	-0.00106	-0.27705	0.13%	1	519

The smoother, Figure 23(b), estimated a linear effect with 1 effective degree freedom (Table 10). This effect was, however, non-significant at the 5% level (p=0.484).

3.1.7 Confirmatory Analysis

3.1.7.1 Modelling the Risk of *S. mansoni* Infection

3.1.7.1.1 Model Selection

The individual effects of water contact site, history of schistosomiasis treatment, frequency of water contact and treatment for soil-transmitted helminths in the past year were significantly associated with the risk of *S. mansoni* infection in the simple logistic regression analysis (Table 11).

The relative effects of these covariates in the multiple logistic regression model are presented in Table 12. Using the deviance-based test, STH treatment was dropped and the resulting model became our main effects model (Table 13).

Table 11: Simple logistic regression models for the risk of *S. mansoni* infections in Alabonu

	Simple Regression Models			Deviance Analysis		
Terms	**OR (95% CI)**	**SE**	**p-value**	**D_0 - D_1**	**df**	**p-value**
Water contact point						
Others	1					
Tutukope	4.33 (1.47, 18.55)	0.625	0.019	8.890	2	0.012
Akpeakpe	2.44 (0.83, 10.41)	0.623	0.153			
Treatment for schistosomiasis						
Never	1					
Ever	0 (NA, 8.12 x 10^{12})	824.921	0.985	5.810	1	0.016
Frequency of water contact						
Daily	1					
1-3 times per week	0.26 (0.04, 0.86)	0.733	0.064	5.090	1	0.024
Sex						
Females	1					
Males	0.91 (0.51, 1.58)	0.287	0.734	0.120	1	0.729
STH treatment in the past year						
No	1					
Yes	0.60 (0.33, 1.06)	0.297	0.088	3.050	1	0.000
STH treatment						
Never	1					
Ever	0.68 (0.39, 1.16)	0.280	0.161	2.000	1	0.157
Schistosomiasis treatment in the past year						
No	1					
Yes	0 (NA, 7.09 x 10^{18})	594.164	0.982	1.490	1	0.222

Table 12: Multiple logistic regression model for *S. mansoni* infections in Alabonu.

	Multiple Regression Model		
Model Terms	**OR (95% CI)**	**SE**	**p-value**
Water contact point			
Others	1		
Tutukope	4.61 (1.55, 19.8)	0.628	0.015
Akpeakpe	2.47 (0.83, 10.1)	0.626	0.149
Treatment for schistosomiasis			
Never	1		
Ever	0 (NA, 9.65 x 10^{14})	797.837	0.985
Frequency of water contact			
Daily	1		
1-3 times per week	0.24 (0.04, 0.82)	0.743	0.055
STH treatment in the past year			
No	1		
Yes	0.63 (0.34, 11.2)	0.303	0.126

Table 13: Deviance analysis for terms in the multiple regression model for *S. mansoni* infections in Alabonu. The second column gives the difference between the null (D_0) and residual (D_1) deviances as each term is added sequentially to the model.

	Deviance Analysis				
Model Terms	***df***	**D_0 - D_1**	**Residual Deviance**	**Residual *df***	**p-value**
Null Model			**371.190**	**516**	
Water contact point					
Others					
Tutukope	2	8.581	362.610	514	0.014
Akpeakpe					
Treatment for schistosomiasis					
Never					
Ever	1	6.646	355.970	513	0.010
Frequency of water contact					
Daily					
1-3 times per week	1	4.834	351.130	512	0.028
STH treatment in the past year					
No					
Yes	1	2.433	348.700	511	0.119

The effects of interactions between the terms in the main effects model are presented in Table 14. None of these terms were, however, statistically significant and the deviance-based test further indicated these terms could be omitted from the model without causing any significant increase in the overall deviance. Therefore, the main effects model became our provisional model for the risk of *S. mansoni* infection.

Table 14: Deviance analysis for interaction terms in the main effects model for *S. mansoni* infections in Alabonu.

	Deviance Analysis				
Model Terms	***df***	**D_0 - D_1**	**Residual Deviance**	**Residual *df***	**p-value**
Main Effects Model			**351.130**	**512.000**	
Water contact point : Schistosomiasis treatment	2	0.000	351.130	510.000	1.000
Water contact point : Frequency of water contact	2	4.679	346.460	510.000	0.096
Schistosomiasis treatment : Frequency of water contact	1	0.000	351.130	511.000	1.000

3.1.7.2 Goodness-of-Fit Check

The dispersion statistic, 0.68, fell within our acceptable cut-off limit, as discussed in section 2.6.2.2.1. We, therefore, ruled out over-dispersion.

3.1.7.3 Model Validation

3.1.7.3.1 Assessment of Homoscedasticity

Figure 24 suggests no obvious violation of the homoscedasticity assumption. Therefore, it would seem at this point that the model may not require any further improvement.

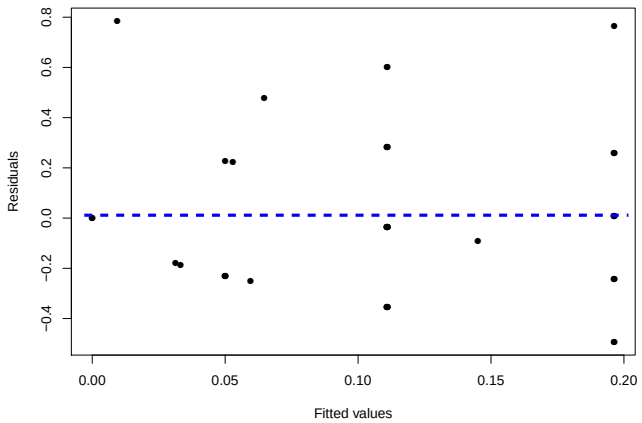


Figure 24: Grouped Pearson residuals versus fitted values of the provisional model for *S. mansoni* infections in Alabonu.

3.1.7.3.2 Assessing the Assumption of Independence

A cursory inspection of Figure 26 mainly seems to suggest the absence of any clear segregation of negative or positive residuals. However, a comparison with the raw response data, Figure 27, also seems to suggest our provisional model may have under-estimated the risk of *S. mansoni* infection in the areas around the centre of the community. To get a better assessment of any latent spatial patterns, an empirical variogram was computed for the residuals (Figure 25).

3.1.7.3.3 Variogram Analysis

The shape of the omni-directional empirical variogram, Figure 25, seems relatively flat with not sill, thereby pointing to the absence of spatial dependence. However, the omni-directional variogram assumes isotropy. We, therefore, proceeded to verify this assumption by computing anisotropic variograms.

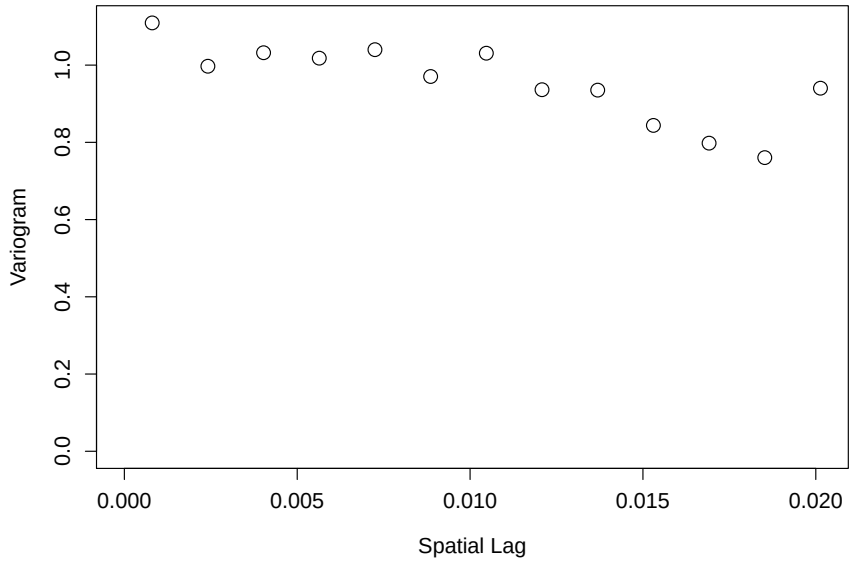


Figure 25: Omni-directional empirical variogram for the Pearson residuals of the provisional model for *S. mansoni* infection in Alabonu.

The anisotropic variograms, Figure 28, for the four directions appear the same, thereby ruling out any obvious violation of the isotropy assumption. Hence, the omni-directional variogram was used in the rest of the assessment. In line with our earlier observations, Figure 29 confirms the absence of spatial autocorrelation. Moreover, our formal test for trend indicated a non-significant effect (p=0.759).

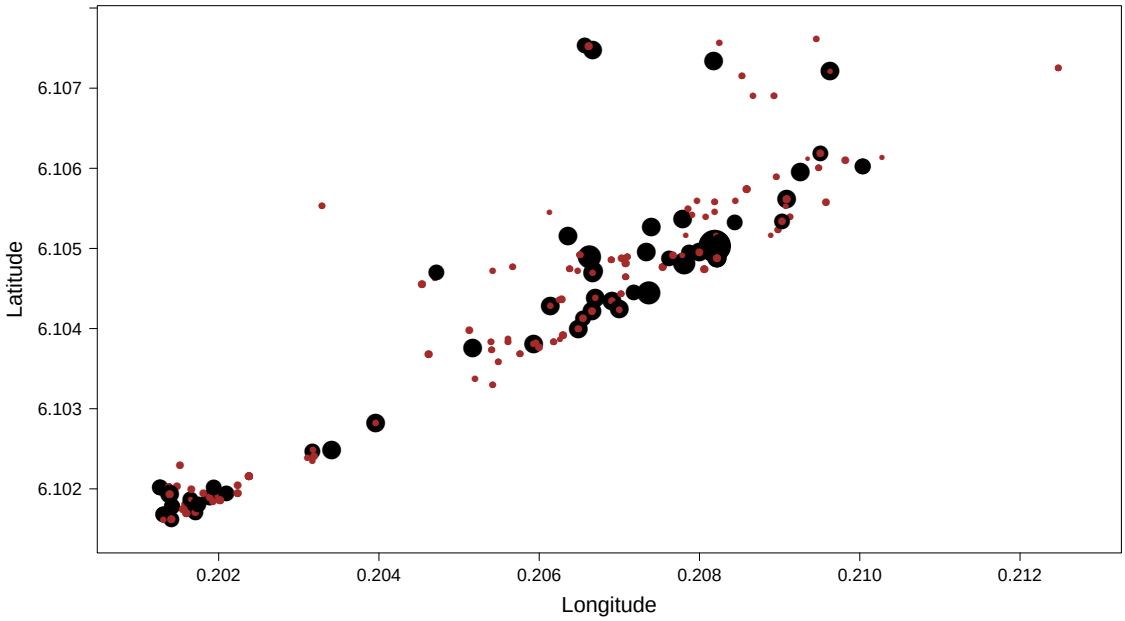


Figure 26: Spatially-referenced residuals of the provisional model for *S. mansoni* infection in Alabonu. The radii of the circles are proportional to the absolute values of the Pearson residuals. The red and black circles represent the negative and positive residuals, respectively.

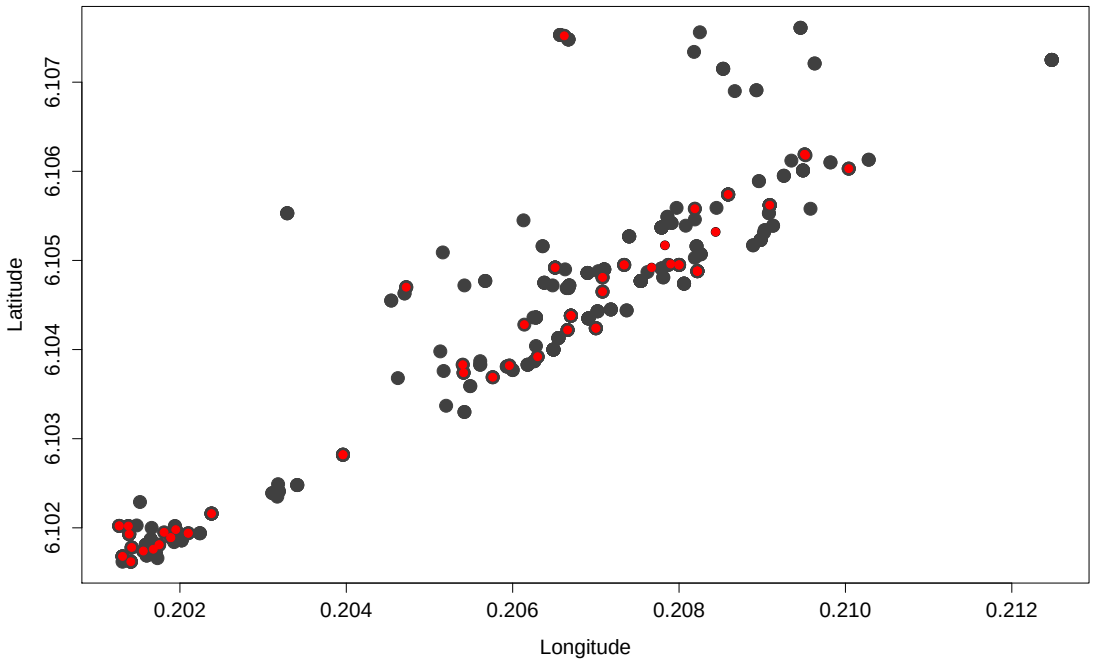


Figure 27: Distribution of *S. mansoni* infection in Alabonu. Each point corresponds to a sampled location i.e location of the household of the sampled participant. The red and black points represent the infected and non-infected cases, respectively.

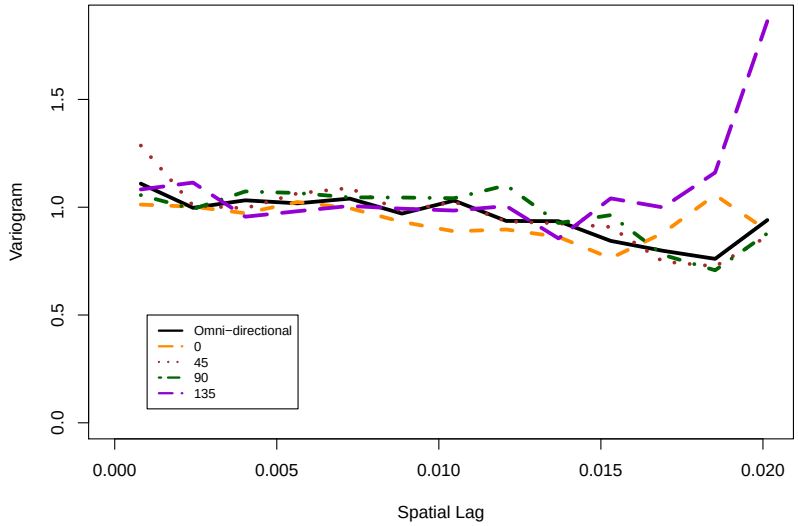


Figure 28: Directional empirical variograms for the 0° , 45°, 90° and 135° spatial directions (computed with a tolerance angle of ± 22.5°).

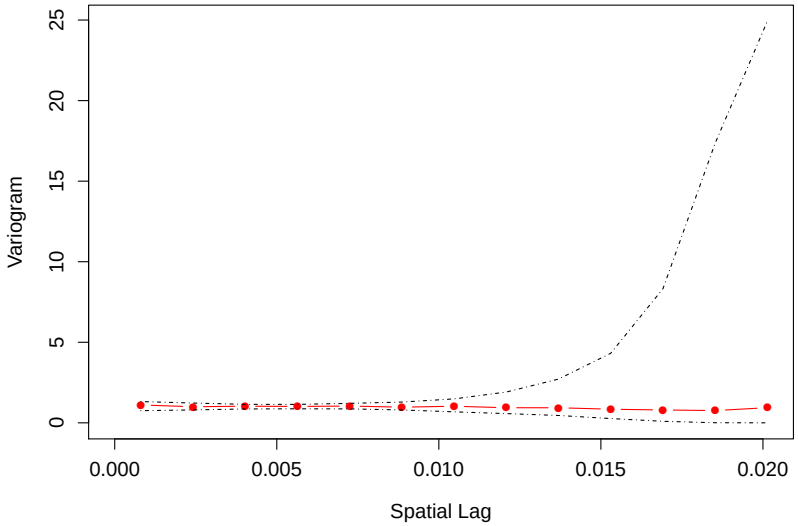


Figure 29: Empirical variogram (dotted red line) for the Pearson residuals of the provisional model for *S. mansoni* infection in Alabonu and the envelope of random permutations.

3.1.7.3.4 Final Model for *S. mansoni* infection

The results, therefore, suggest the community-level variation in the risk of *S. mansoni* infection was adequately explained by the water contact sites' frequented by the inhabitants as well as their history of schistosomiasis treatment and frequency of exposure (equation 20).

$$\begin{aligned} \text{logit}(\pi_i) \quad = \quad & 0.05 \quad + \quad 4.63 \quad (\textbf{Tutukope}) \quad + \\ & 2.37 \quad (\textbf{Akpeakpe}) \quad + \quad 0.00 \quad (\textbf{treatment}) \quad + \\ & 0.26 \quad (\textbf{Frequency of water contact}) \end{aligned} \tag{20}$$

3.2 Klamadaboe

This community was within a 1 mile distance from Torgorme, the most developed of the three communities under consideration in this study. There was also a footpath linking the two communities (Figure 30), hence the inhabitants of Klamadaboe had the choice of walking over the 1 mile distance to collect treated water from the communal tap in Torgorme. But what factors would influence people's decision to choose tap water over the Lake for domestic purposes? Or would anyone walk over that distance just to collect a bucketful of water when the Lake was practically in their backyards? Our study questionnaire for this community was, therefore, designed to collect exposure data in the context of the inhabitants' level of schistosomiasis awareness (Appendix A).

Figure 30: Left: A section of Klamadaboe. Right: The footpath that linked Klamadaboe to the neighbouring community, Torgorme.

The footpath also provided an alternative means of transportation by motorbike taxis. The community, however, lacked public conveniences and a school, hence children from this community either attended school in Torgorme or in communities across the Lake. Therefore, even though Klamadaboe had no history mass chemotherapy intervention, we made a reasonable assumption at the beginning of the study that the children from this community who attended school in neighbouring Torgorme or other places were more likely to have received anthelmintic treatment at some point in time during any of the school-based mass chemotherapy intervention programmes.

3.2.1 Baseline Characteristics of Study Population

Ten (7.87%) of the study population were infected with *S. haematobium* whilst *S. mansoni* infections occurred in 23 (17.69%). Co-infection with both schistosome species was recorded in 3 (2.34%) of the study participants. The *S. haematobium* infections occurred in 2(4.28%) preschool-aged children, 4 (11%) school-aged and 4 (6%) adults.

The *S. mansoni* infections were, however, observed in 17 (29.82%) of adults, 5(14.28 %) school-aged children and 1 (6.66 %) preschool-aged child. None of the children in the preschool-aged category was co-infected with *S. haematobium* and *S. mansoni*. The prevalence of light and intensity heavy *S. haematobium* infections in the community were 8 (6%) and 2 (1.5%), respectively. The heavily infected participants occurred in both the school-aged and adult categories. All the *S. mansoni* infections were, however, light intensity infections.

3.2.1.1 Individual-Level Treatment History

Despite our initial assumption that most children in this community may have received treatment, the questionnaire survey revealed that most of them had never indeed been treated and that only 1.75% of the school-aged children in the community had ever received treatment from school (Table 15). The treatment history was also equally low among adults in the community.

3.2.1.2 General Awareness of Schistosomiasis

The level of schistosomiasis awareness was generally low across all age groups in the community (Table 16). However, school-aged children had the lowest level of awareness. For instance, all girls in the school-aged group were not aware of the most common symptoms of acute schistosomiasis such as haematuria, dysuria and bloody diarrhoea. The highest proportion of inhabitants who had experienced any of the common symptoms of schistosomiasis were boys in the school-aged category.

3.2.1.3 Water Contact Activities

Almost all the inhabitants were exposed to the Lake on daily basis (Table 17). Of the activities that involved contact with Lake, water collection, washing, swimming and canoe boarding were the most common whilst sand gathering from the bottom of the Lake as well as fishing and wading through parts of the Lake were the less common activities.

Moreover, whilst swimming was more common among school-aged children in the community, contact activities among adult females mainly involved water collection and washing (Table 17).

3.2.2 Micro-Level Determinants of Infectivity

As a pre-intervention site where human exposure patterns were virtually constant, it was only logical to assume human infection levels. However, the levels of infectivity associated with different exposure events may also be influenced by the type of exposure activities as well as the duration of exposure. Exposure activities including swimming and washing often tend to involve long contact with the Lake, and hence may be associated with higher risk of infection. Moreover, the manner in which such activities were performed may also have varied across the community. The rest of this section, therefore, examines some of the potential micro-level factors that may have interacted with the measured covariates to influence risk in the community.

Empirical evidence suggests the chemicals in soap could have adverse effects the aquatic larval stages of the schistosome parasites (Okwuosa & Osuala 1993). Therefore, depending on the manner in which activities that involved the use of soap were carried out, washing and bathing may may be associated with lower levels of risk. For instance, if the inhabitants were in the habit of standing in the shallow margins of the Lake while they washed, then the potential effect of the lather on the schistosome larvae could have indeed rendered washing a low risk activity. However, if washing was carried out on the shore instead, then the toxic effect of soap on the schistosome larvae would have been negligible.

Table 15: Schistosomiasis treatment history of the inhabitants of Klamadaboe. The numbers in brackets represent the row percentages.

			Management of Symptoms of Schistosomiasis Experienced				No. of Praziquantel Treatment			Period Since Last Treatment		Where Treatment was Administered			
Sex	**Age-Group**	**Total**	**Did Nothing**	**Bought Medicine**	**Hospital**	**Used Herbs**	**Once**	**Twice**	**Never**	**A Year Ago**	**> 2 years Ago**	**Hospital**	**VRA**	**Bought**	**School**
Females	5-15 years	24	4 (16.67)	0 (0.00)	0 (0.00)	0 (0.00)	1 (4.17)	0 (0.00)	21 (87.50)	1 (4.17)	0 (0.00)	0 (0.00)	0 (0.00)	1 (4.17)	0 (0.00)
	16-30 years	26	2 (7.69)	0 (0.00)	2 (7.69)	0 (0.00)	0 (0.00)	2 (7.69)	21 (80.77)	0 (0.00)	2 (7.69)	0 (0.00)	0 (0.00)	1 (3.85)	1 (3.85)
	31-40 years	10	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (10.00)	1 (10.00)	0 (0.00)	0 (0.00)	2 (20.00)	0 (0.00)	1 (10.00)	1 (10.00)	0 (0.00)
	41-60 years	9	2 (22.22)	0 (0.00)	1 (11.11)	0 (0.00)	1 (11.11)	0 (0.00)	5 (55.56)	0 (0.00)	1 (11.11)	1 (11.11)	0 (0.00)	0 (0.00)	0 (0.00)
Males	5-15 years	33	14 (42.42)	2 (6.06)	1 (3.03)	0 (0.00)	4 (12.12)	1 (3.03)	23 (69.70)	3 (9.09)	2 (6.06)	1 (3.03)	1 (3.03)	2 (6.06)	1 (3.03)
	16-30 years	15	5 (33.33)	0 (0.00)	2 (13.33)	0 (0.00)	5 (33.33)	0 (0.00)	5 (33.33)	0 (0.00)	5 (33.33)	2 (13.33)	1 (6.67)	1 (6.67)	1 (6.67)
	31-40 years	5	1 (20.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (40.00)	0 (0.00)	1 (20)	0 (0.00)	2 (40.00)	1 (20.00)	1 (20.00)	0 (0.00)	0 (0.00)
	41-60 years	13	2 (15.38)	2 (15.38)	1 (7.69)	0 (0.00)	5 (38.46)	0 (0.00)	5 (38.46)	0 (0.00)	5 (38.46)	1 (7.69)	2 (15.38)	1 (7.69)	0 (0.00)

Table 16: Local knowledge of schistosomiasis in Klamadaboe. The numbers in brackets represent the row percentages.

			Awareness of Schistosomiasis			Knowledge of Common Symptoms				Symptoms of Schistosomiasis Experienced		
Sex	**Age-Group**	**Total**	**Very Well**	**Not Very Well**	**No**	**Responses**	**Haematuria**	**Dysuria**	**Bloody Diarrhoea**	**Haematuria**	**Dysuria**	**Bloody Diarrhoea**
Females	5-15 years	24	0 (0.00)	2 (8.33)	20 (83.33)	No idea	22 (91.67)	22 (91.67)	22 (91.67)	2 (8.33)	2 (8.33)	2 (8.33)
						True	0 (0.00)	0 (0.00)	0 (0.00)			
						False	0 (0.00)	0 (0.00)	0 (0.00)			
	16-30 years	26	6 (23.08)	10 (38.46)	7 (26.92)	No idea	7 (26.92)	9 (34.62)	10 (38.46)	3 (11.54)	3 (11.54)	2 (7.69)
						True	16 (61.54)	13 (50.00)	10 (38.46)			
						False	0 (0.00)	1 (3.85)	3 (11.54)			
	31-40 years	10	0 (0.00)	7 (70.00)	2 (20.00)	No idea	3 (30.00)	4 (40.00)	6 (60.00)	0 (0.00)	0 (0.00)	0 (0.00)
						True	6 (60.00)	5 (50.00)	3 (30.00)			
						False	0 (0.00)	0 (0.00)	0 (0.00)			
	41-60 years	9	2 (22.22)	3 (33.33)	1 (11.11)	No idea	1 (11.11)	2 (22.22)	2 (22.22)	0 (0.00)	0 (0.00)	3 (33.33)
						True	5 (55.56)	4 (44.44)	4 (44.44)			
						False	0 (0.00)	0 (0.00)	0 (0.00)			
Males	5-15 years	33	3 (9.09)	10 (30.30)	16 (48.48)	No idea	19 (57.58)	18 (54.55)	21 (63.64)	14 (42.42)	12 (36.36)	7 (21.21)
						True	10 (30.30)	11 (33.33)	6 (18.18)			
						False	0 (0.00)	0 (0.00)	2 (6.06)			
	16-30 years	15	2 (13.33)	8 (53.33)	1 (6.67)	No idea	2 (13.33)	4 (26.67)	5 (33.33)	6 (40.00)	5 (33.33)	6 (40.00)
						True	9 (60.00)	6 (40.00)	6 (40.00)			
						False	0 (0.00)	1 (6.67)	0 (0.00)			
	31-40 years	5	1 (20.00)	2 (40.00)	0 (0.00)	No idea	0 (0.00)	0 (0.00)	1 (20.00)	1 (20.00)	1 (20.00)	0 (0.00)
						True	3 (60.00)	3 (60.00)	0 (0.00)			
						False	0 (0.00)	0 (0.00)	2 (40.00)			
	41-60 years	13	2 (15.38)	9 (69.23)	0 (0.00)	No idea	1 (7.69)	2 (15.38)	6 (46.15)	3 (23.08)	4 (30.77)	3 (23.08)
						True	10 (76.92)	8 (61.54)	5 (38.46)			
						False	0 (0.00)	1 (7.69)	0 (0.00)			

Table 17: Activities that took the inhabitants of Klamadaboe into contact with the Lake. The numbers in brackets represent the row percentages.

			Frequency of Water Contact			Water Contact Activity						
Sex	**Age-Group**	**Total**	**1-3pwk**	**1 pmth**	**Daily**	**Water Collection**	**Fording**	**Washing**	**Swimming**	**Fishing**	**Canoe Boarding**	**Sand Winning**
Females	5-15 years	24	1 (4.16)	0 (0.00)	21 (87.50)	17 (70.83)	0 (0.00)	8 (33.33)	19 (79.17)	0 (0.00)	9 (37.50)	0 (0.00)
	16-30 years	26	0 (0.00)	0 (0.00)	23 (88.46)	23 (88.46)	1 (3.85)	16 (61.54)	3 (11.54)	1 (3.85)	12 (46.15)	8 (30.77)
	31-40 years	10	0 (0.00)	0 (0.00)	9 (90.00)	8 (80.00)	0 (0.00)	4 (40.00)	1 (10.00)	1 (10.00)	3 (30.00)	3 (30.00)
	41-60 years	9	1 (11.11)	0 (0.00)	5 (55.56)	5 (55.56)	0 (0.00)	2 (22.22)	2 (22.22)	0 (0.00)	2 (22.22)	0 (0.00)
Males	5-15 years	33	3 (9.09)	0 (0.00)	26 (78.79)	21 (63.64)	0 (0.00)	12 (36.36)	25 (75.76)	2 (6.06)	10 (30.30)	0 (0.00)
	16-30 years	15	1 (6.67)	0 (0.00)	10 (66.67)	6 (40.00)	0 (0.00)	4 (26.67)	8 (53.33)	3 (20.00)	2 (13.33)	3 (20.00)
	31-40 years	5	1 (20.00)	0 (0.00)	2 (40.00)	3 (60.00)	0 (0.00)	2 (40.00)	1 (20.00)	1 (20.00)	2 (40.00)	1 (20.00)
	41-60 years	13	0 (0.00)	0 (0.00)	11 (84.62)	7 (53.85)	3 (23.08)	3 (23.08)	6 (46.15)	6 (46.15)	4 (30.77)	1 (7.69)

Moreover, some households in the community had "bathing huts" (Figure 31), thereby implying that not everyone bathed inside the Lake or the shore. Therefore, the factors that influenced the use of these "bathing huts" rather than the Lake may have had important implications on the variation in risk. For instance, did the age and sex of the members of any given household in the community determine the choice of where they bathed?

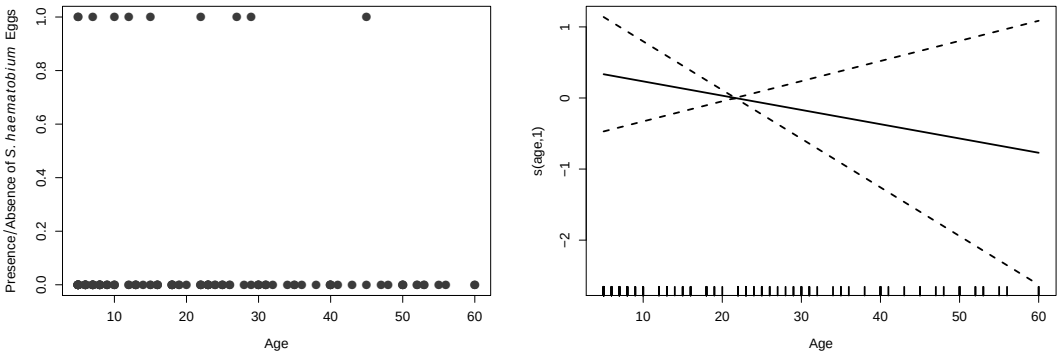
Figure 31: A hut, used as a bathing area (circled), in Klamadaboe. Adults in households that had these bathing huts were assumed to use them instead of bathing directly inside the littoral zone of the Lake.

3.2.3 Exploratory Analysis

3.2.3.1 Assessing Non-Linearity in Age

3.2.3.1.1 *S. haematobium* Infection

Figure 32(a) indicates almost all the observed *S. haematobium* infections in the community occurred among participants aged ≤ 30 years whilst the older participants virtually harboured no infections. In line with this observation, the non-parametric smoother, Figure 32(b), estimated a decreasing effect with age. However, this estimated linear effect was not significant at the 5% level, p=0.406 (Table 18).



(a) Scatterplot (b) Smoother

Figure 32: (a) Scatterplot of the prevalence of *S. haematobium* infection in Klamadaboe versus age of the study participants. (b) Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines).

3.2.3.1.2 *S. mansoni* Infection

An initial assessment of the relationship between age and the prevalence of *S. mansoni* infection suggested that though infections were common among participants aged ≤ 30 years, most of these infections were mainly concentrated among those in the 20 - 30 years age range (Figure 33). The non-parametric smoother consequently estimated a non-linear effect in age, Figure 34(a), with 2.019 effective degrees of freedom as the

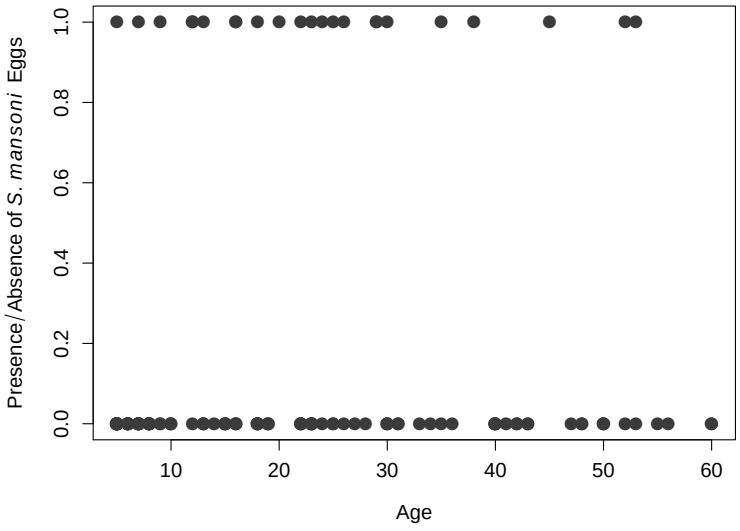
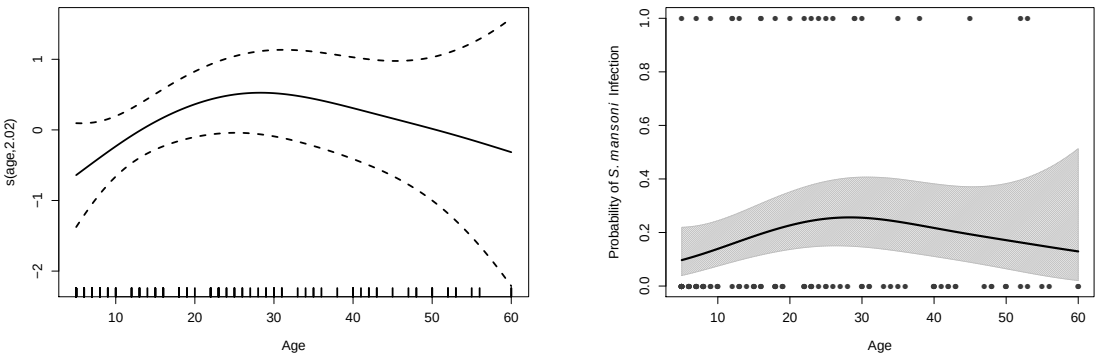


Figure 33: Scatterplot of the prevalence of *S. mansoni* infection in Klamadaboe versus age of the study participants.

Table 18: Model summary for the smoothing function of age in equation 1 for Klamadaboe. The smoothers were estimated using default settings in the R mgcv package by Wood 2006. Therefore, the optimum amount of smoothing, EDF, was determined at default by cross-validation.

	Approximate Significance of Smooth Terms					Fit Details				
Outcome	**Smoother Term**	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance explained**	**Scale est.**	**N**
S. haematobium	s (Age)	1	1	0.689	0.406	-0.00306	-0.4231	1.08%	1	127
S. mansoni	s (Age)	2.019	2.519	4.25	0.172	0.0228	-0.060777	4.35%	1	130

optimum amount of smoothing (Table 18). The highest odds of infection was estimated as 0.24 and this occurred around the age of 25 (Figure 34(b)). However, the p-value for the smoother, p=0.172, indicates the effect was non-significant at the 5% level (Table 18).



(a) Scatterplot

(b) Smoother

Figure 34: (a) Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines). (b) Estimated non-parametric smoother (solid line) for age, plotted on the scale of the raw data. The grey dots represent the observed prevalence of *S. mansoni* infection in Klamadaboe.

3.2.4 Confirmatory Analysis

3.2.4.1 Modelling the Risk of *S. haematobium* Infection

3.2.4.2 Model Selection

The simple logistic regression models for *S. haematobium* infection are presented in Appendix B. None of the measured covariates, however, contributed significantly to explaining the risk of *S. haematobium* infection in the community.

3.2.5 Modelling the Risk of *S. mansoni* Infection

3.2.5.1 Model Selection

The effects of the measured covariates in the simple logistic regression models are presented in Table 19. Local awareness of schistosomiasis and history of praziquantel treatment were found to be significantly associated with the risk of *S. mansoni* infection.

Table 19: Simple regression models for the risk of *S. mansoni* infections in Klamadaboe

		Simple Regression Models			Deviance Analysis		
Model Terms	**No. of Participants**	**OR (95% CI)**	**SE**	**p-value**	**D_0-D_1**	***df***	**p-value**
Schistosomiasis Awareness							
Yes, very well	16	1					
Yes, but not very well	51	1.49	0.723	0.584	7.843	2	0.020
No	47	0.27	0.879	0.139			
History of Praziquantel Treatment							
Once	19	1					
Twice or more	4	0.43 (0.02, 4.23)	1.260	0.501	7.492	2	0.024
Never	87	0.19 (0.06, 0.62)	0.599	0.005			
Source of Water for Domestic Use							
Tap	2	1					
Lake	107	3.28×10^6 (0, NA)	2.40×10^6	0.995	0.752	2	0.687
Both tap and Lake	1	1 (0, 4.21×10^{41})	3.39×10^6	1.000			
Sex							
Females	69	1					
Males	66	1.15 (0.47, 2.88)	0.460	0.756	0.100	1	0.752
Last Praziquantel Treatment							
Up to a year ago	4						
Over 2 years ago	19	1.40 (0.11, 33.66)	1.320	0.799	0.066	1	0.797
Frequency of Contact with the Lake							
1-3 times per week	7	1					
1 per month	0						
Daily	107	1.05 (0.16, 20.69)	1.126	0.968	0.000	1	1.000
Activities that involved contact with the Lake:							
Swimming							
No	49	1					
Yes	65	0.48 (0.17, 1.29)	0.512	0.147	2.137	1	0.144
Sand Winning							
No	98	1					
Yes	16	1.94	0.648	0.307	0.970	1	0.325
Water Collection							
No	24	1					
Yes	90	0.69 (0.23, 2.36)	0.5837	0.526	0.390	1	0.532
Fording							
No	110	1					
Yes	4	1.63 (0.08, 13.57)	1.183	0.680	0.160	1	0.689
Canoe Boarding							
No	70						
Yes	44	1.22 (0.43, 3.31)	0.513	0.699	0.150	1	0.699
Fishing							
No	100	1					
Yes	14	0.77 (0.11, 3.19)	0.809	0.752	0.100	1	0.752
Washing							
No	63	1					
Yes	51	1.10 (0.4, 2.97)	0.505	0.854	0.030	1	0.862

The relative effects of of schistosomiasis awareness and praziquantel treatment history are presented in (Table 20). The retention or omission of any of these terms from the model was justified by the deviance-based test, Table 21. Schistosomiasis awareness ended up becoming the only term in our provisional model for *S. mansoni* infections in Klamadaboe.

Table 20: Multiple regression model for *S. mansoni* infections in Klamadaboe

Terms	OR (95% CI)	SE	p-value
Schistosomiasis Awareness			
Yes, very well	1		
Yes, but not very well	1.55 (0.38, 8.07)	0.756	0.561
No	0.40 (0.06, 2.73)	0.939	0.330
History of Praziquantel Treatment			
Once	1		
Twice or more	0.40 (0.02, 3.96)	1.266	0.464
Never	0.28 (0.08, 0.99)	0.634	0.047

Table 21: Deviance Test for terms in the multiple regression model for *S. mansoni* infections in Klamadaboe. The second column represents the difference between the null (D_0) and residual (D_1) deviance as terms are added sequentially, starting from first to last

	Deviance Analysis				
Terms	*df*	D_0 - D_1	Residual Deviance	Residual *df*	p-value
Schistosomiasis Awareness					
Yes, very well					
Yes, but not very well	2	7.67	92.025	103	0.022
No					
History of Praziquantel Treatment					
Once					
Twice or more	2	3.90	88.127	101	0.142
Never					

3.2.5.2 Goodness-of-Fit Check

The value of the Pearson dispersion statistic, 0.87, occurred within our acceptable range. This was, therefore, used as the basis for ruling out over-dispersion in the provisional model.

3.2.5.3 Model Validation

3.2.5.3.1 Assessing the Homoscedasticity Assumption

The interpretation of any trend in the spread of the grouped Pearson residuals across the range of the fitted values in Figure 35 is rendered difficult by the small sample size of 135. However, there does not seem to be any obvious violation of the homoscedasticity assumption either.

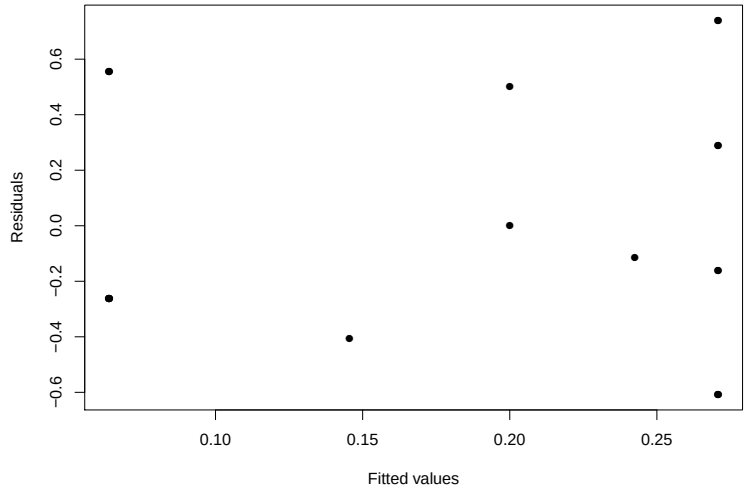


Figure 35: Grouped Pearson residuals versus fitted values of the provisional model for *S. mansoni* infections in Klamadaboe. The small sample size makes the interpretation of any trend in the spread of the residuals difficult.

3.2.5.3.2 Assessing the Assumption of Independence

A visual assessment of Figure 36 mainly suggests the absence of any apparent clusters of identical residuals, except for a small area where the likelihood of infection was increased (the circled area in Figure 36). This area, with an increased likelihood of risk, corresponds to the part of the community where our provisional model over-estimated the risk of *S. mansoni* infection (Figure 37).

However, our interpretation of any latent spatial trend in Figure 36 may be subjective. In the next section, we employ the variogram as a more objective tool in assessing spatial dependence, or lack of it, in the point-referenced residuals.

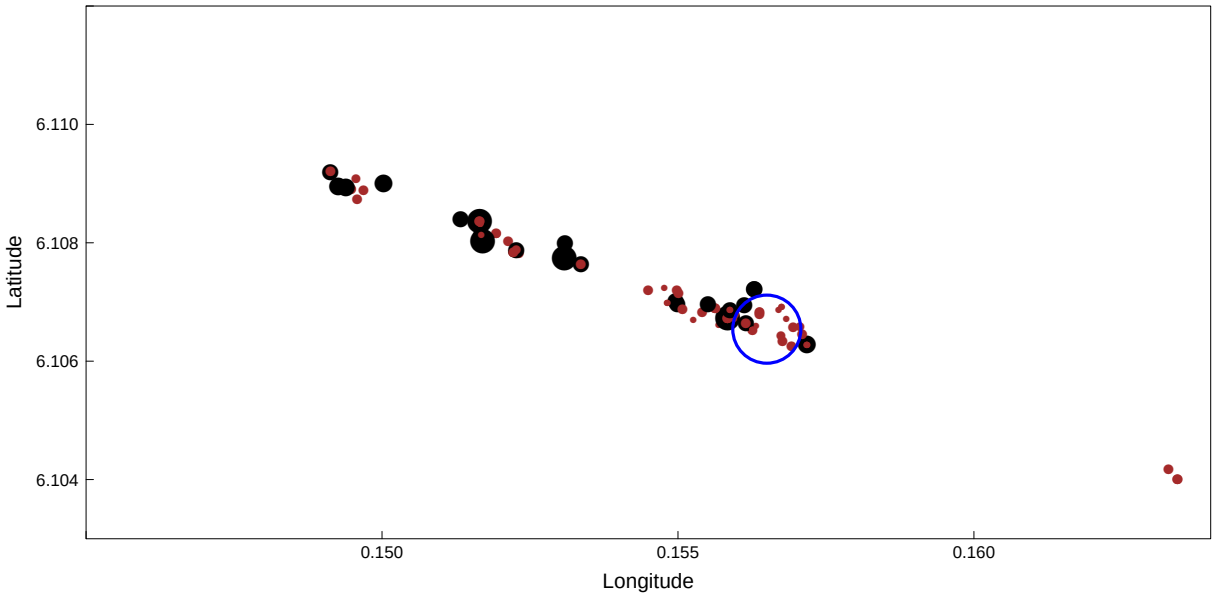


Figure 36: Spatially-referenced residuals of the provisional model for *S. mansoni* infection in Klamadaboe. The radii of the circles are proportional to the absolute values of the Pearson residuals. The red and black circles represent the negative and positive residuals, respectively.

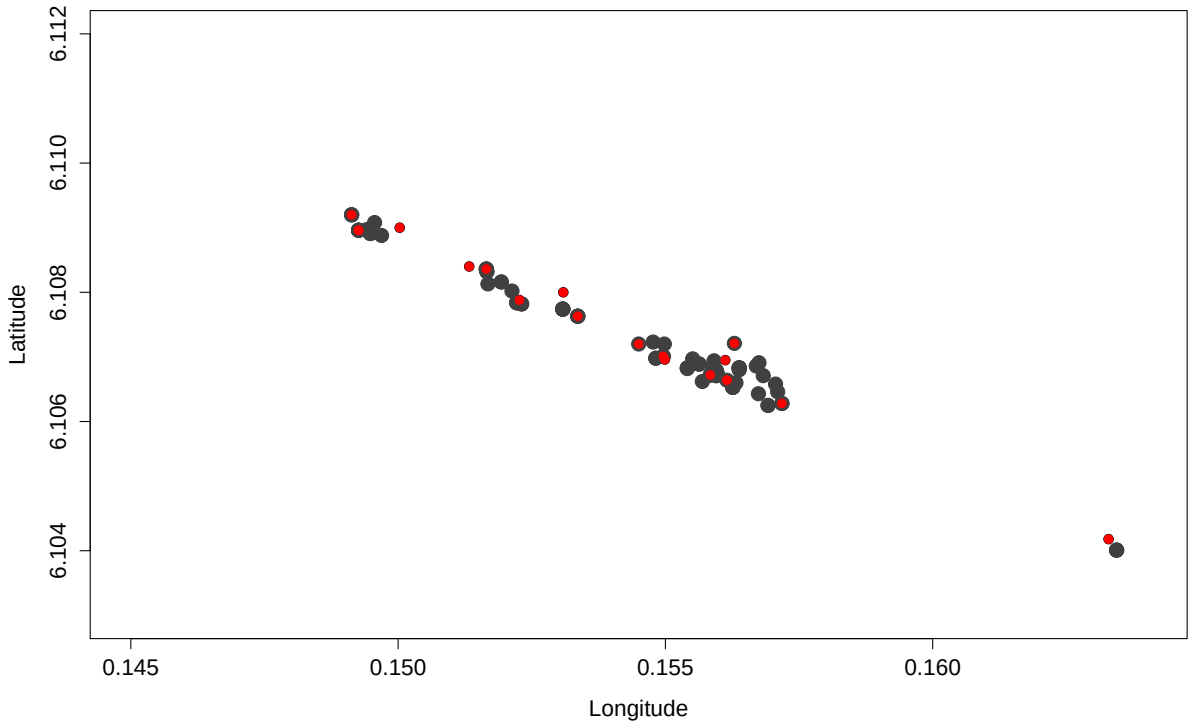


Figure 37: Distribution of *S. mansoni* infection in Klamadaboe. Each point corresponds to a sampled location i.e location of the household of the sampled participant. The red and black points represent infected and non-infected cases, respectively.

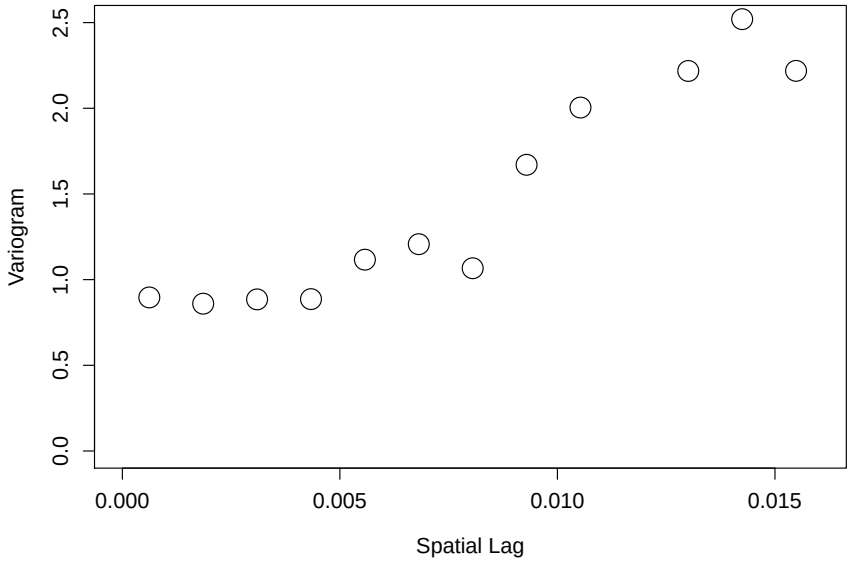


Figure 38: Omni-directional empirical variogram for the Pearson residuals of the provisional model for *S. mansoni* infection in Klamadaboe.

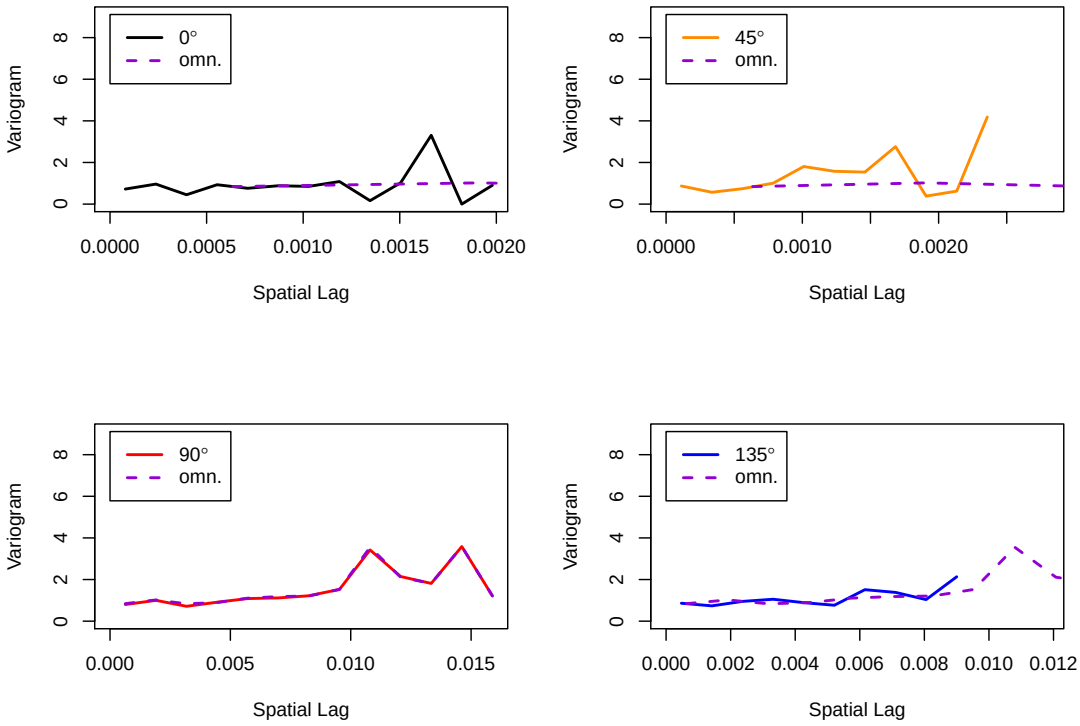


Figure 39: Directional empirical variograms for lags in the 0°, 45°, 90° and 135° spatial directions, computed with a tolerance angle of ± 22.5°.

3.2.5.3.3 Variogram Analysis

The omni-directional empirical variogram, Figure 38, suggests an increasing spatial trend till a range of 0.013°, which may point to the presence of spatial dependence. Our assessment for directional dependence, however, indicated that the isotropy assumption may be reasonable in this particular case (Figure 39). Therefore, the omni-directional variogram was regarded as the mean variogram for all spatial directions and used in the rest of the assessment. As judged by the envelope of random permutations (Figure 40) and the formal test for trend, $p = 0.248$, there was no statistically significant spatial dependence in the residuals.

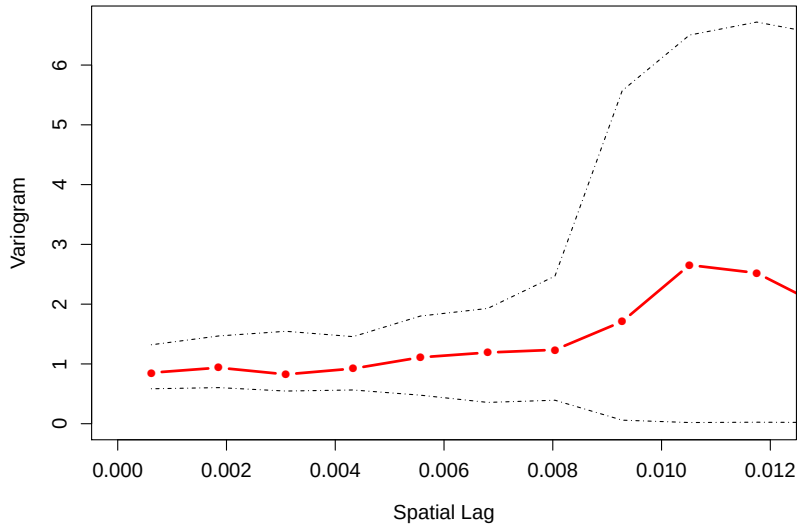


Figure 40: The omni-directional empirical variogram of the Pearson residuals of the provisional model for *S. mansoni* infection in Klamadaboe (dotted red line) and the envelope of random permutations (outer dashed lines).

3.2.5.4 Final Model for *S. mansoni* Infections

Therefore, having ruled out the relevance of the unmeasured micro-level factors, equation 21 became our final model for *S. mansoni* infection.

$$\begin{aligned} \text{logit}(\pi_i) &= 0.25 + 1.49 \quad (\textbf{awareness}_{notverywell}) \\ &+ 0.27 \quad (\textbf{awareness}_{no}) \end{aligned} \tag{21}$$

3.3 Torgorme

This was the most developed of the three communities under consideration in this study. It had a health centre, an access road, public sanitation facilities and a communal tap (Figure 41). Moreover, the community was in receipt of sporadic chemotherapy interventions by the Volta River Authority. We, therefore, regarded Torgorme at the time of recruitment as a post-intervention site with residual transmissions. Given the access road and the communal tap, inhabitants of this community had the choice of using alternative means of transportation, other than canoes, as well as treated water.

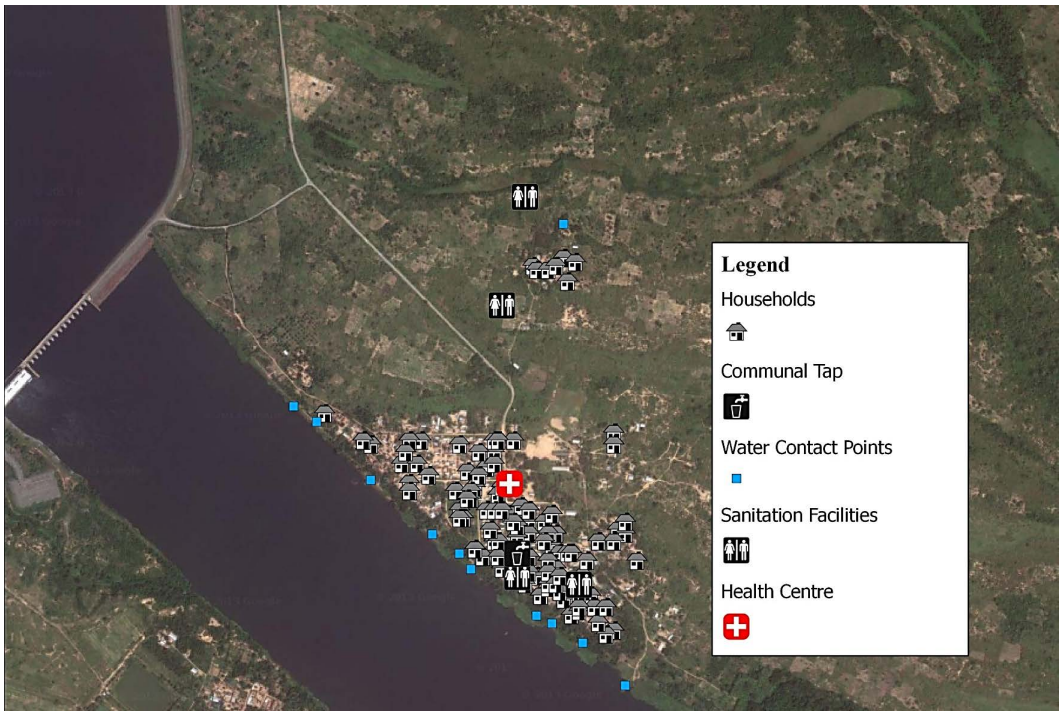


Figure 41: Map of Torgorme showing its proximity to the Akosombo Dam; and the distribution of water contact sites and social amenities in the community.

The communal ownership of the tap also meant that inhabitants had unrestricted access, thereby ruling out any differential access to treated water as a possible source of variation in risk at the community level. However, the supply of potable water was also interrupted on weekly basis, hence the presence of the tap did not completely rule out the Lake as a source of water for domestic use, even among those inhabitants who may have preferred treated water over the Lake. The questionnaire for this community

was, therefore, designed mainly to collect exposure data in the context of the kind of activities that took inhabitants into contact with the Lake. Moreover, we sought to determine if the chemotherapy intervention programmes had been accompanied by health education, which would have meant that the inhabitants would have had higher levels of schistosomiasis awareness.

3.3.1 Baseline Characteristics of the Study Population

S. haematobium infections occurred in 10 (3.70%) of the study population whilst *S. mansoni* infections were observed in 2 (0.72%). The *S. mansoni* and 7(2.59%) of the *S. haematobium* infected cases were school-aged children. All the observed *S. haematobium* and *S. mansoni* infections in the community were light intensity infections.

3.3.1.1 Individual-Level Treatment History

Even though we initially regarded Torgorme as a post-intervention site, the questionnaire survey indicated that 62.34% of the participants had never received treatment for schistosomiasis (Table 22). Nineteen percent of our study participants had, however, been treated on at least one occasion. Of the 124 people who reported experiencing symptoms of acute schistosomiasis at some point in time, 56% of them did nothing to treat the symptoms.

3.3.1.2 General Awareness of Schistosomiasis

The proportion of our study participants who reported being unaware of schistosomiasis included 56% of the school-aged population, 13% of the young adults, 23% of the people between the ages of 31 and 40; and 22% of the above 40 age group (Table 23). Symptoms of acute schistosomiasis were most common among boys in the school-aged population.

3.3.1.3 Water Contact Activities

The frequency of exposure to the Lake varied between daily, thrice weekly and monthly among individuals of all age groups in the community (Table 24). The most common reasons for going to the Lake were for water collection and canoe boarding while the less common activities included sand winning from the bottom of the Lake, fishing and fording through parts of the Lake on the way to the farm. Moreover, swimming was a common activity among the school-aged children in the community.

3.3.2 Micro-Level Determinants of Infectivity

In this section, we examine some of the more obvious micro-level factors that may have potentially acted in conjunction with the measured covariates to influence the transmission dynamics of *Schistosoma* infections in Torgorme. The role of these micro-level factors are examined under environmental, socio-economic and behavioural factors.

3.3.2.1 Environmental Factors

The lentic conditions of the Lake, that made it a suitable habitat for the intermediate snail host of the schistosome species, is one of the direct consequences of the impoundment of the Volta River in creating the Akosombo Dam. The spillways of the Dam are, however, opened during the major rainy season, which occurs between June and August, to release excess water. This action generally increases the current of the Lake downstream. However, since Torgorme is situated immediately after Dam (Figure 42(a)), the current tends to be stronger due to the higher flow volume (Figure 42(b)).

Therefore, there could be considerable reductions in the population density of the snail hosts during this period; as most of them are likely to be carried downstream with the current. Moreover, the food supply of these snails is also known to be swept away by such strong current (Cowper 1971). These variations in snail population density may, therefore, have important implications on the overall transmission dynamics of schistosome infections in the community. For instance, the *Bulinus spp* that serve as intermediate snail hosts for *S. haematobium* are known to cope better and establish themselves in running water than the snail hosts for *S. mansoni* (Cowper 1971).

Table 22: Schistosomiasis treatment history of the inhabitants of Torgorme. The numbers in brackets represent the row percentages.

			Management of Symptoms of Schistosomiasis Experienced				No. of Praziquantel Treatment			Period Since Last Treatment		Where Treatment was Administered			
Sex	**Age-Group**	**Total**	**Did Nothing**	**Bought Medicine**	**Hospital**	**Used Herbs**	**Once**	**Twice**	**Never**	**A Year Ago**	**> 2 years Ago**	**Hospital**	**VRA***	**Bought**	**School**
Females	5-15 years	88	21 (23.86)	1 (1.14)	1 (1.14)	0 (0.00)	8 (9.09)	1 (1.14)	65 (73.86)	5 (5.68)	4 (4.55)	1 (1.14)	3 (3.41)	1 (1.14)	5 (5.68)
	16-30 years	39	8 (20.51)	0 (0.00)	2 (5.13)	2 (5.13)	12 (30.77)	1 (2.56)	19 (48.72)	8 (20.51)	5 (12.82)	2 (5.13)	7 (17.95)	2 (5.13)	2 (5.13)
	31-40 years	35	6 (17.14)	0 (0.00)	2 (5.71)	1 (2.86)	4 (11.43)	0 (0.00)	23 (65.71)	0 (0.00)	4 (11.43)	3 (8.57)	1 (2.86)	0 (0.00)	0 (0.00)
	41-60 years	43	4 (9.30)	3 (6.97)	3 (6.97)	0 (0.00)	11 (25.58)	0 (0.00)	25 (58.14)	1 (2.33)	10 (23.26)	2 (4.65)	9 (20.93)	0 (0.00)	0 (0.00)
Males	5-15 years	83	27 (32.53)	4 (4.82)	5 (6.02)	1 (1.20)	14 (16.87)	1 (1.20)	56 (67.47)	10 (12.05)	5 (6.02)	5 (6.02)	0 (0.00)	3 (3.61)	7 (8.43)
	16-30 years	13	2 (15.38)	2 (15.38)	3 (23.08)	1 (7.69)	8 (61.54)	1 (7.69)	2 (15.38)	4 (30.77)	5 (38.46)	2 (15.38)	3 (23.08)	1 (7.69)	3 (23.08)
	31-40 years	4	0 (0.00)	0 (0.00)	1 (25.00)	0 (0.00)	2 (50.00)	1 (25)	0 (0.00)	0 (0.00)	3 (75.00)	0 (0.00)	2 (50.00)	0 (0.00)	1 (25.00)
	41-60 years	11	1 (9.09)	2 (18.18)	2 (18.18)	0 (0.00)	1 (9.09)	2 (18.18)	7 (63.64)	0 (0.00)	3 (27.27)	0 (0.00)	2 (18.18)	1 (9.09)	0 (0.00)

*VRA: Volta River Authority

Table 23: Local knowledge of schistosomiasis in Torgorme. The numbers in brackets represent the row percentages.

			Awareness of Schistosomiasis			Knowledge of Common Symptoms				Symptoms of Schistosomiasis Experienced		
Sex	**Age-Group**	**Total**	**Very Well**	**Not Very Well**	**No**	**Responses**	**Haematuria**	**Dysuria**	**Bloody Diarrhoea**	**Haematuria**	**Dysuria**	**Bloody Diarrhoea**
Females	5-15 years	88	15 (17.05)	12 (13.64)	47 (53.41)	No idea True False	39 (44.32) 30 (34.09) 5 (5.68)	43 (48.86) 24 (27.27) 7 (7.95)	44 (50.00) 21 (23.86) 9 (10.23)	8 (9.09)	13 (14.77)	10 (11.36)
	16-30 years	39	11 (28.21)	15 (38.46)	6 (15.38)	No idea True False	6 (15.38) 23 (58.97) 3 (7.69)	10 (25.64) 18 (46.15) 4 (10.26)	16 (41.03) 7 (17.95) 9 (23.08)	3 (7.69)	8 (20.51)	7 (17.95)
	31-40 years	35	7 (20.00)	11 (31.43)	9 (25.71)	No idea True False	7 (20.00) 18 (51.43) 2 (5.71)	13 (37.14) 10 (28.57) 4 (11.43)	17 (48.57) 3 (8.57) 7 (20.00)	0 (0.00)	5 (14.29)	5 (14.29)
	41-60 years	43	6 (13.95)	20 (46.51)	11 (25.58)	No idea True False	10 (23.26) 27 (62.79) 0 (0.00)	14 (32.56) 19 (44.19) 4 (9.30)	24 (55.81) 4 (9.30) 9 (20.93)	5 (11.63)	6 (13.95)	2 (4.65)
Males	5-15 years	83	6 (7.23)	17 (20.48)	48 (57.83)	No idea True False	46 (55.42) 21 (25.30) 4 (4.82)	48 (57.83) 19 (22.89) 4 (4.82)	53 (63.86) 12 (14.46) 6 (7.23)	23 (27.71)	31 (37.35)	18 (21.69)
	16-30 years	13	1 (7.69)	9 (69.23)	1 (7.69)	No idea True False	1 (7.69) 9 (69.23) 1 (7.69)	2 (15.38) 8 (61.54) 1 (7.69)	4 (30.76) 5 (38.46) 2 (15.38)	5 (38.46)	3 (23.08)	5 (38.46)
	31-40 years	4	0 (0.00)	3 (75.00)	0 (0.00)	No idea True False	0 (0.00) 3 (75) 0 (0.00)	1 (25.00) 2 (50.00) 0 (0.00)	2 (50.00) 1 (25.00) 0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	41-60 years	11	3 (27.27)	6 (54.55)	1 (9.09)	No idea True False	2 (18.18) 8 (72.73) 0 (0.00)	4 (36.36) 6 (54.55) 0 (0.00)	8 (72.73) 2 (18.18) 0 (0.00)	3 (27.27)	3 (27.27)	2 (18.18)

Table 24: Activities that took the inhabitants of Torgorme into contact with the Lake. The numbers in brackets represent the row percentages.

			Frequency of Water Contact			Water Contact Activity						
Sex	**Age-Group**	**Total**	**1-3pwk**	**1 pmth**	**Daily**	**Water Collection**	**Fording**	**Washing**	**Swimming**	**Fishing**	**Canoe Boarding**	**Sand Winning**
Females	5-15 years	88	38 (43.18)	16 (18.18)	20 (22.73)	59 (67.05)	0 (0.00)	23 (26.14)	34 (38.64)	1 (1.14)	28 (31.82)	3 (3.41)
	16-30 years	39	18 (46.15)	13 (33.33)	1 (2.56)	29 (74.36)	1 (2.56)	13 (33.33)	4 (10.26)	0 (0.00)	21 (53.85)	4 (10.26)
	31-40 years	35	11 (31.43)	11 (31.43)	4 (11.43)	22 (62.86)	0 (0.00)	9 (25.71)	2 (5.71)	0 (0.00)	9 (25.71)	4 (11.43)
	41-60 years	43	10 (23.26)	23 (53.49)	4 (9.30)	31 (72.09)	1 (2.33)	11 (25.58)	3 (6.98)	1 (2.33)	13 (30.23)	2 (4.65)
Males	5-15 years	83	29 (34.94)	14 (16.87)	28 (33.73)	50 (60.24)	3 (3.61)	17 (20.48)	45 (54.22)	9 (10.84)	21 (25.30)	2 (2.41)
	16-30 years	13	2 (15.38)	5 (38.46)	4 (30.77)	6 (46.15)	0 (0.00)	0 (0.00)	4 (30.77)	4 (30.77)	4 (30.77)	2 (15.38)
	31-40 years	4	0 (0.00)	1 (25.00)	2 (50.00)	2 (50.00)	1 (25.00)	3 (75.00)	1 (25.00)	1 (25.00)	2 (50.00)	0 (0.00)
	41-60 years	11	2 (18.18)	5 (45.45)	3 (27.27)	5 (45.45)	0 (0.00)	3 (27.27)	0 (0.00)	1 (9.09)	5 (45.45)	1 (9.09)

(a) (b)

(c) (d)

Figure 42: Figure 42(a): Spillways of the Akosombo Dam as seen from the shoreline in Torgorme. Figure 42(b): The opened spillways (source: (Wikipedia 2014)). Figure 42(c): The low flow rate at a water contact site when the spillways are closed. Figure 42(d): The increased current when the spillways are opened during the raining season.

3.3.2.2 Socio-Economic and Behavioural Factors

None of the households in the community relied solely on Lake water for domestic purposes. Therefore, as far as the source of water in the home was concerned, households in the community could be categorised into two: those that used tap water alone and those that used water from both the Lake and the communal tap. Seventy-one percent of the households in the community occurred in the former group. Bearing in mind that there were regular interruptions in the flow of tap water, the micro-level factors that determined why households fell in any of these categories could have had important implications on the transmission dynamics of infections in the community.

Figure 43: Earthenware pots that were used for water storage in some households in Torgorme.

For instance, many households in the community used earthenware pots for water storage (Figure 43). However, these pots were usually not big enough to hold the quantity of water that would sustain any household for more than a few days. Therefore, access to larger water storage containers, such as barrels, could have been a possible reason why some households managed to use tap water even during the periods when flow was interrupted. Proximity of households to the Lake or tap could have also been a contributory factor. People who lived further away from the tap and closer to the Lake may have drunk from the tap but relied on the Lake for domestic purposes.

The time of day and manner in which water contact activities were carried out could also have had implications on infectivity. The washing of clothes was normally performed in the late mornings or early afternoons. But were the inhabitants in the habit of washing their clothes directly in the littoral zone of the Lake as shown in Figure 44? If that was indeed the case, then the toxicity of the soap may have had lethal effects on the infective larval stages of the schistosomes. Moreover, were most women in the habit of taking their pre-school aged children with them to Lake (Figure 44). If that was a common practice in the community, then the contaminating habit of these children would play an important role in the transmission cycle of *Schistosoma* infections.

Figure 44: A woman (circled on the left) washing clothes at one of the water contact points in Torgorme. Standing nearby is her pre-school aged son (circled on the right).

Even though the inhabitants had access to alternative means of transportation, canoe transport was still relatively common (Table 24). Therefore, the micro-level factors that determined the choice of canoes over taxis or motorbike taxis could help in understanding the transmission dynamics in the community. Such micro-level factors may have included the kind of goods that were being transported across the Lake as well as the canoe charges, as compared to taxi fares.

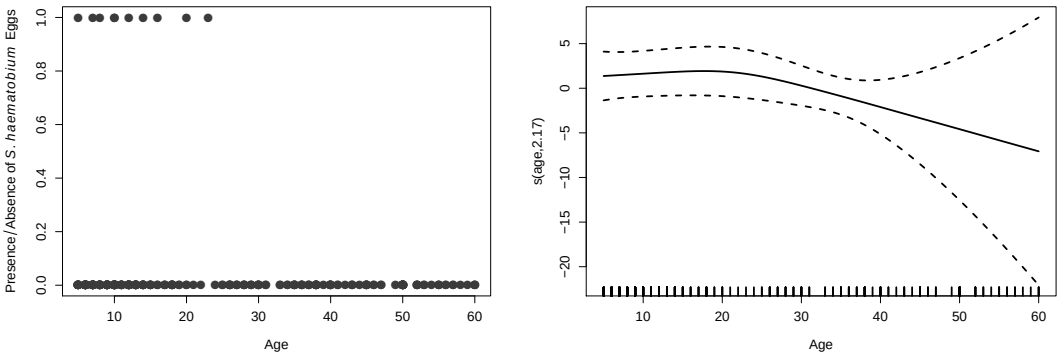
3.3.3 Exploratory Analysis

3.3.3.1 Assessing Non-Linearity in Age

3.3.3.2 *S. haematobium* Infection

An assessment of the relationship between age and the prevalence of *S. haematobium* infection indicated that all the observed infections occurred in participants aged ≤ 25 years (Figure 45(a)). The non-parametric smoother, which was employed in discerning the nature of the age-prevalence relationship, estimated a non-linear effect with 2.172 effective degrees of freedom (Figure 45(b)). In effect, the smoothing curve suggested an almost constant trend till the age of 25, followed by a steady decline across the remaining age range. The wide confidence intervals for the ≥ 40 years trend conformed with the non-detection of infections in that age group. The overall estimated effect of age was

non-significant at the 5% level, p=0.479 (Table 25).

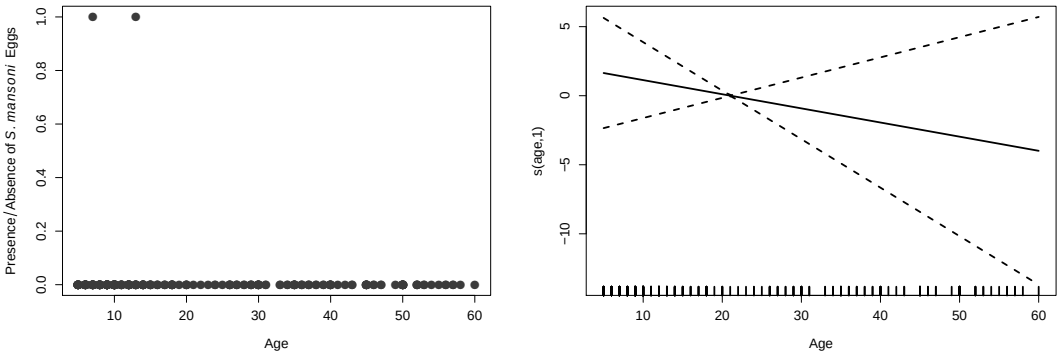


(a) Scatterplot (b) Smoother

Figure 45: (a) Scatterplot of the prevalence of *S. haematobium* infection in Torgorme versus age of the study participants. (b) Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines).

3.3.3.3 *S. mansoni* Infection

Figure 46(a) indicates the two observed *S. mansoni* infections occurred in children. Consequently, the smoother estimated a non-significant linear effect with a decreasing trend across the age range for the sampled population (Figure 46(b) and Table 25).



(a) Scatterplot (b) Smoother

Figure 46: (a) Scatterplot of the prevalence of *S. mansoni* infection in Torgorme versus age of the study participants. (b) Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines).

Table 25: Model summary for the smooth function of age in the generalised additive model (equation 1).

	Approximate Significance of Smooth Terms					Fit Details				
Outcomes	**Smooth Terms**	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance expl.**	**Scale est.**	**N**
S. haematobium	s (Age)	2.172	2.739	2.182	0.479	0.0166	-0.69334	10.60%	1	270
S. mansoni	s (Age)	1	1	0.678	0.410	-0.00046	-0.90491	6.47%	1	275

3.3.4 Confirmatory Analysis

Due to the very low prevalence of *S. mansoni* infections in the community, we chose to focus the formal analysis on *S. haematobium* infection.

3.3.4.1 Modelling the Risk of *S. haematobium* Infection

3.3.4.2 Model Selection

The effects of the measured covariates in the simple logistic regression models for *S. haematobium* infection are presented in Table 26. With the exception of age1, which was marginally significantly associated with *S. haematobium* infection, none of the other covariates had a significant effect.

Table 26: Simple regression models for *S. haematobium* infections in Torgorme. D_0 - D_1 represents the difference between the null and residual deviance.

		Simple Regression Models			Deviance Analysis		
Model Terms	**No. of Participants**	**OR (95% CI)**	**SE**	**p-value**	**D_0 - D_1**	***df***	**p-value**
Age							
age1		0.93 (0.85, 1.00)	0.038	0.071		1	
age2		0.00 (NA, 9.94 x 10^{43})	1424	0.992		1	
History of Praziquantel Treatment							
Once	60	1					
Twice or more	7	1.00 (0, Inf)	7.15 x 10^5	1.000	4.229	2	0.121
Never	197	3.72 x 10^5 (0, NA)	2.48 x 10^6	0.994			
Frequency of Contact with the Lake							
1-3 times per week	110	1					
1 per month	88	0.32 (0.02, 2.22)	1.129	0.314	1.248	2	0.536
Daily	66	0.80 (0.11, 4.23)	0.882	0.798			
Source of Water for Domestic Use							
Tap	188	1					
Lake	0						
Both tap and Lake	76	0.43 (0.02, 2.60)	1.090	0.441	0.713	1	0.398
Schistosomiasis Awareness							
Yes, very well	49	1					
Yes, but not very well	93	0.93 (0.09, 20.33)	1.240	0.952	0.467	2	0.792
No	123	1.61 (0.23, 31.99)	1.134	0.675			
Sex							
Females	205	1					
Males	111	0.81 (0.17, 2.99)	0.702	0.763	0.093	1	0.760
Last Praziquantel Treatment							
Up to a year ago	28						
Over 2 years ago	39	-	9.56 x 10^6	1.000	-3.36 x 10^{-10}	1	1.000
Activities that involved contact with Lake:							
Sand Winning							
No	247	1					
Yes	18	0 (NA, 4.18 x 10^{60})	1.630 x 10^3	0.993	1.04	1	0.308
Fishing							
No	248	1					
Yes	17	0 (NA, 6.80 x 10^{62})	1.684 x 10^3	0.993	0.972	1	0.324
Washing							
No	186	1					
Yes	79	1.75 (0.34, 8.15)	0.778	0.472	0.499	1	0.480
Fording							
No	259	1					
Yes	6	0.00 (NA, 1.34 x 10^{60})	1.615 x 10^3	0.993	0.381	1	0.537
Water Collection							
No	61	1					
Yes	204	1.72 (0.28, 32.88)	1.092	0.620	0.277	1	0.599
Swimming							
No	172	1					
Yes	93	1.45 (0.28, 6.74)	0.777	0.632	0.223	1	0.637
Canoe Boarding							
No	162	1					
Yes	103	1.10 (0.21, 5.12)	0.776	0.899	0.016	1	0.899

In line on our discussion in section 3.1.3.1.4, 35 years was chosen as the break-point for the manual programming of Figure 45(b).

3.3.4.3 Goodness-of-Fit Check

The value, 0.30, for the dispersion statistic of the provisional model fitted within our cut-off limit as specified in section 2.6.2.2.1. Based on this, we concluded the model was not over-dispersion.

3.3.4.4 Model Validation

3.3.4.4.1 Assessing the Homoscedasticity Assumption

The spread of residuals across the range of the fitted values in Figure 47 is consistent with our binary outcome data.

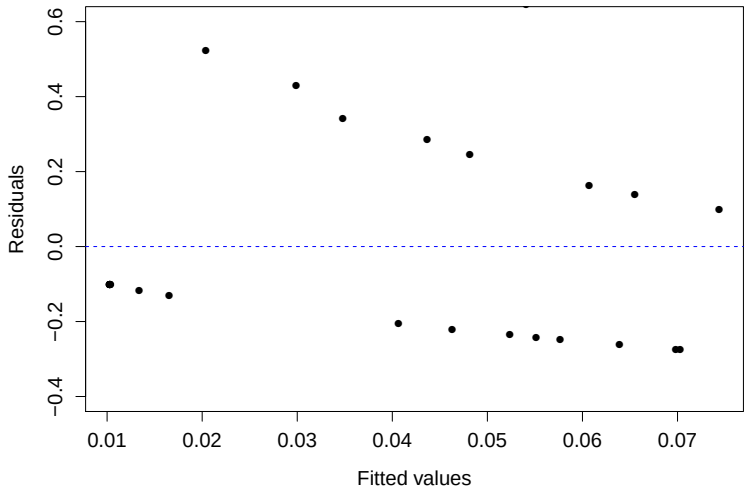


Figure 47: Grouped Pearson residuals versus fitted values of the provisional model for *S. haematobium* infections in Torgorme.

3.3.4.4.2 Assessing the Assumption of Independence

The point map for the residuals, Figure 48, shows some clusterings of positive residuals which correspond to the areas where our provisional model seemed to have under-estimated the risk of *S. haematobium* infection (Figure 49). However, the model mainly over-estimated the risk of *S. haematobium* infection across the community (Figure 49). To form a better judgement of any present latent spatial trend, we employed the use of the variogram in quantifying the spatial dependence.

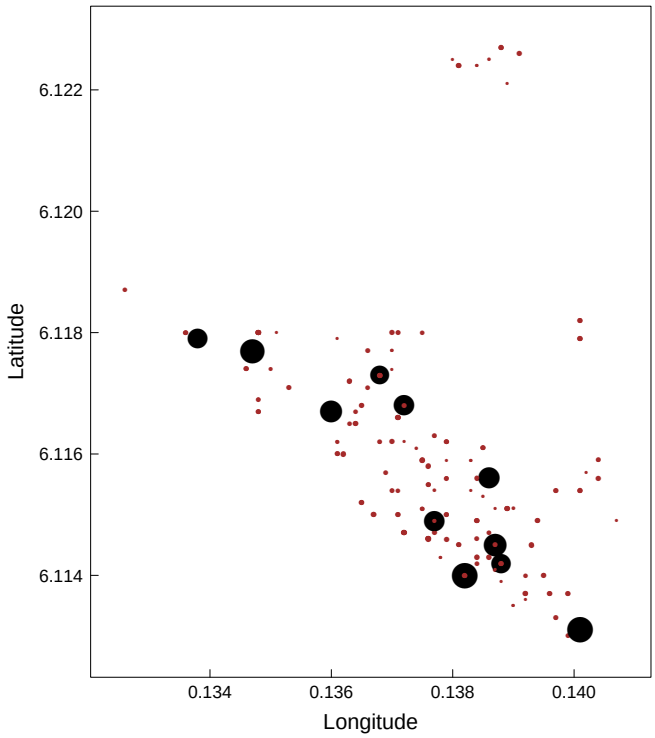


Figure 48: Spatially-referenced residuals of the provisional model for *S. haematobium* infection in Torgorme. The radii of the points are proportional to the absolute values of the Pearson residuals. The red and black points represent the negative and positive residuals, respectively.

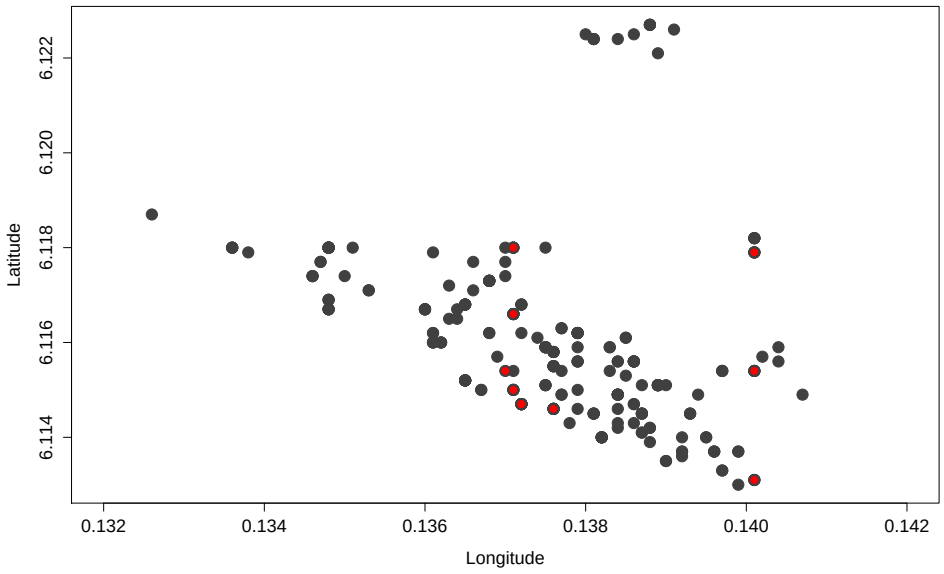


Figure 49: Distribution of *S. haematobium* infection in Alabonu. Each point corresponds to a sampled location i.e location of the household of the sampled participant. The red and black points represent infected and non-infected cases, respectively.

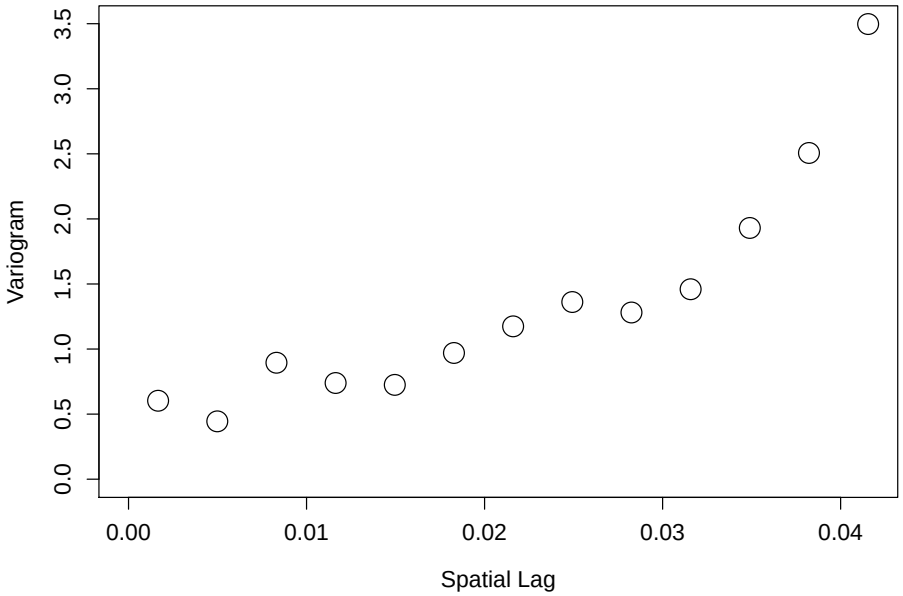


Figure 50: Omni-directional empirical variogram for the Pearson residuals of the provisional model for *S. haematobium* infection in Torgorme.

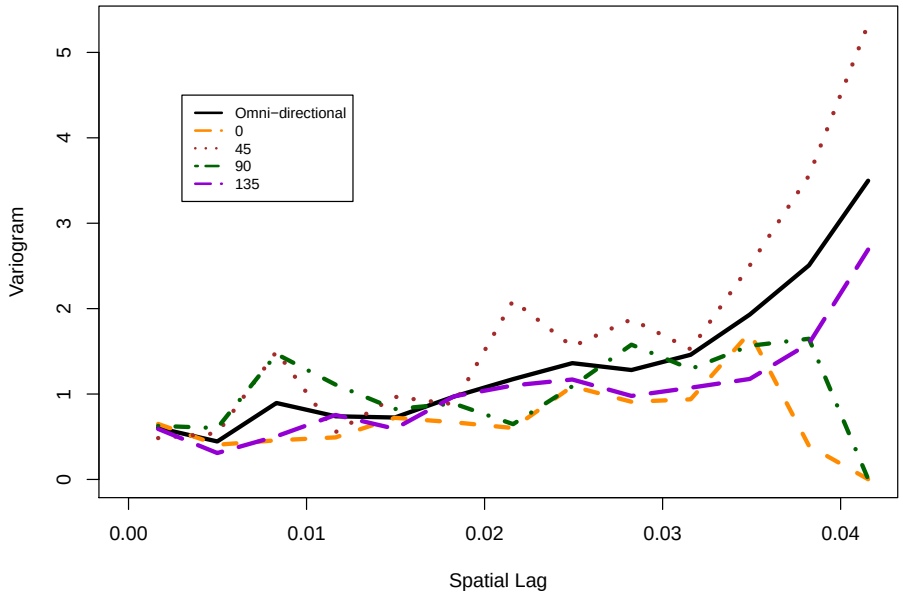


Figure 51: Directional empirical variograms for the 0° , 45°, 90° and 135° spatial directions (with a tolerance angle of ± 22.5° around each of these directions).

3.3.4.4.3 Variogram Analysis

The shape of the omni-directional empirical variogram, Figure 50, is consistent with an unbounded variogram. This, therefore, seems to suggest that while the spatial process might be intrinsic, there may not be any spatial correlation (Webster & Oliver 2001).

Judging by the slopes of the anisotropic variograms, the isotropy assumption seemed reasonable. Therefore, the omni-directional variogram was regarded as a mean variogram for all spatial directions and used in the rest of the analysis. Both the envelope of random permutations, Figure 52, and our formal test for trend indicated the absence of any significant spatial autocorrelation (p =0.486).

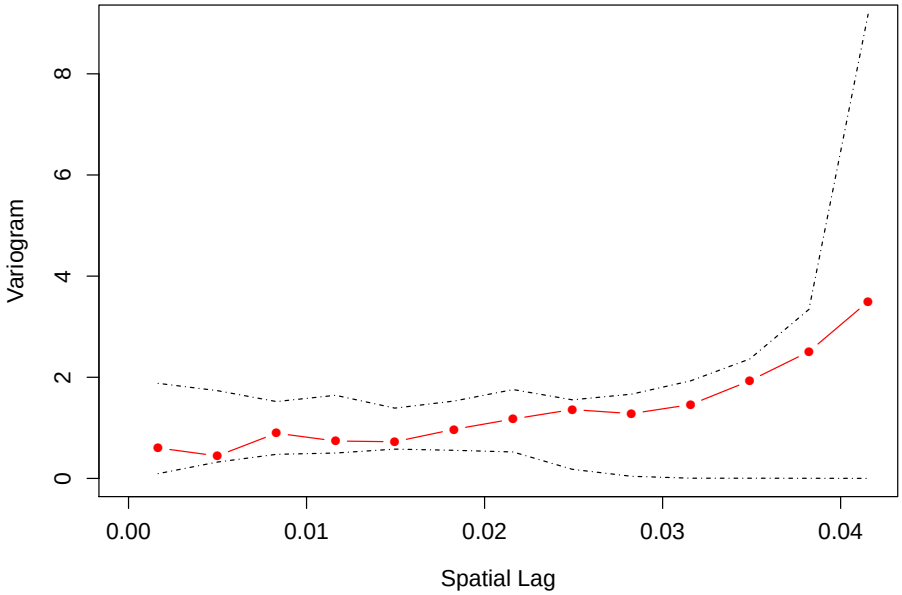


Figure 52: Empirical variogram for the Pearson residuals of the provisional model for *S. haematobium* infections in Torgorme.

3.3.4.5 Final Model for *S. haematobium* Infection

Based on these results, we ruled out the presence of any statistically significant heterogeneities in the distribution of risk across the community. Therefore, equation 22 became our final model for *S. haematobium* infection.

$$\text{logit}(\pi_i) = 0.11 + 0.93 \quad (\textbf{age1}) \tag{22}$$

4 Discussion

The transmission dynamics of the schistosome species and the variation in risk at the local level are known to be influenced by complex interactions between macro-level factors such as local environmental conditions and a series of micro-level factors resulting mainly from the geographical relation between the inhabitants of such endemic settings and the water bodies that harbour the intermediate snail hosts of the schistosome species. Since the micro-level factors that influence human exposure patterns may differ for different endemic settings, the patterns of schistosome transmission also tend to vary accordingly.

One of the proposed explanations for why these micro-level factors may influence exposure patterns, and hence variations in the pattern of transmission at the local level, is the fact that the population density of the infective stages of the schistosome parasites, the cercariae, tends to exhibit diurnal rhythms in their freshwater habitats. Studies have shown that peak densities of these parasites may occur between noon and 15:00 GMT (Farooq & Mallah 1966). Therefore, even though human exposure patterns in endemic settings may occur at random, the likelihood of acquiring infections may automatically increase for exposure events that occur within the period of maximum cercarial density in the water.

Therefore, the overall transmission potential of infection within any given endemic setting may also be influenced by the time and manner in which exposure activities are performed. Hence, it follows that within settings where economic activities dictate the patterns of transmission, different occupational activities may be associated with different risk of infection.

The present study has investigated the setting-specific variations in the patterns of schistosome transmission in different endemic shoreline communities of the Volta Lake. The different levels of development of the studied communities as well as their treatment histories enabled us to investigate these variations in risk in scenarios where the factors that influenced exposure patterns could be examined from the perspective of inhabitants who had no realistic alternatives to being exposed to the Lake on daily basis, as

opposed to those who may have had other options but still needed the Lake for economic activities.

The presence of significant residual spatial effects would have pointed to the focalisation of risk or spatial autocorrelation and consequently, significant heterogeneities in the patterns of transmission. The implication of such varying likelihood of risk would, therefore, have meant that the selective targeting of chemotherapy interventions at clusters of high risk individuals would present a much more effective approach of controlling transmissions.

However, there was no evidence of significant heterogeneities in the distribution of risk for either *S. haematobium* or *S. mansoni* infections. Possible reasons for these results may include the fact that even though the micro-level factors may have acted in conjunction with the measured covariates, their effects occurred in groups of individuals who did not necessarily live in close proximities to each other. For instance, the farmers and children who experienced exposures within the time of maximum cercarial load lived in different parts of the community. Therefore, our assessment for residual spatial effects, which is based on spatial autocorrelation, may fail when identical risk are not clustered.

The fact that the highest prevalence of *S. haematobium* infections were observed among the school-aged population could be interpreted as lending credence to the current global control strategy, that uses this population in establishing modalities for chemotherapy interventions. However, judging by the fact that we used community-level data, that included both school enrolled and non-enrolled children, we interpret this particular finding with a bit of caution. Our interpretation would have been a lot easier if enough information had been collected in the field to enable us to distinguish between enrolled and non-enrolled children in the studied communities. Empirical evidence from different endemic settings has consistently shown that non-enrolled children tend to harbour more infections due to the longer amount of time they spend in the infested water bodies (Husein et al. 1996, Talaat & Evans 2000).

We also interpret these findings in the context of possible limitations with regard to the diagnostic technique that was used in detecting infection status. Though urine and stool microscopy are widely used as indirect morbidity markers in *Schistosoma-*

endemic settings, single examinations of samples may not yield accurate estimates of infection status, especially in post-intervention settings (Rabello 1997, Richter 2003, WHO 2005).

Therefore, we cannot rule out the possibility of having recorded some false negatives especially in the post-intervention site, Torgorme; in which case the community-level prevalence may have been under-estimated. Moreover, some of the more severe pathologies associated with *S. mansoni* infections are known to influence egg excretion rates. For instance, infected cases with pipe stem fibrosis may excrete very little or no schistosome eggs at all despite harbouring high schistosome burdens. Therefore, the chances of misdiagnosis by microscopy is invariably increased by such factors (de Vlas & Gryseels 1992).

Despite these limitations, however, detection of infections by microscopy has a high level of specificity and is also convenient in field settings for detecting *Schistosoma* infections (Richter 2003, WHO 2005). Moreover, the fact that our samples for *S. haematobium* detection were collected at times of the day when egg excretion is known to be maximum among infected human hosts could be a strength of this study (Doehring et al. 1985).

4.1 Conclusions and Recommendations

The factors that influence the distribution of *Schistosoma* infections in endemic areas are known to vary with the geographical scale of assessment (Brooker 2007). Therefore, while climatic and environmental factors such as proximity to freshwater bodies may influence the risk of infections at large geographical scales (Peng et al. 2010); a series of behavioural and social factors may act in conjunction with ecological factors to influence the local level variation in risk (Woolhouse et al. 1998, Bruun & Aagaard-Hansen 2008).

Moreover, the focalisation of risk which tends to characterise the distribution of infected cases is influenced by local micro-level factors. Therefore, the exact extent of the spatial correlation between events may vary for different endemic settings (Brooker 2007). Detection of the exact extent of these heterogeneities in transmission within different endemic settings could, therefore, be utilised in the effective targeting of chemotherapy interventions.

We have, therefore, investigated heterogeneities in the distribution of risk across different endemic shoreline settings with different levels of development. However, our findings suggest the absence of any significant clustering of risk in the studied communities. Assumptions that went into our models included the fact that micro-level factors specific to each of the studied communities interacted with the measured covariates to influence the transmission of infections. We further assumed that exposure activities involving the micro-level factors occurred at the group level, rather than as discrete activities among individual inhabitants.

While our findings seem to suggest the absence of any statistically significant varying likelihood of risk in the studied communities, we can only interpret these findings in the context of the strength of the covariate effects as well as the amount of variability in the data. Though the micro-level factors that were considered in this study may have acted at the group level, these groups may not have necessarily lived in close proximity to each other. Moreover, regardless of the convergence of the fitting algorithm of the model, the decomposition into large-scale and small-scale variations may not always be reliable (Waller & Gotway 2004). The prevailing question, therefore, is how weak the

effect of the unobserved micro-level factors really were? Moreover, does the absence of any statistically significant spatial autocorrelation really signify that the effect of these setting-specific micro-level factors were not scientifically relevant in explaining the variation in the risk of *Schistosoma* infections in the studied communities?

Though the assessment of the varying likelihood of risk at the local level may provide a means of optimising the effectiveness of chemotherapy implementation in endemic areas, such assessments also tend to be tedious and overly expensive. Indeed, one of the main reasons why the current control strategy uses the district level as its unit of implementation is due to practicality.

Therefore, while our proposed approach may improve the effectiveness of implementation, it would also in effect take control programmes back to square one where the pre-treatment screening of entire populations would have to be somehow circumvented. However, the development of local level model-based risk maps that predict the varying extends of the spatial correlation between events across different endemic settings could provide a solution to avoiding the screening of entire populations.

The main challenge in developing any such risk maps would, therefore, centre around the ability to use selected variables that would transcend the setting-specific micro-level determinants of infectivity. Moreover, since most of the micro-level factors that influence exposure patterns are only best studied qualitatively, the predictive ability of any local level risk maps could be improved if geostatistical methodologies could be extended to the analysis of qualitative data.

References

Arlinghaus, S. L., ed. (1996), *Practical Handbook of Spatial Statistics* , CRC Press.

Bosompem, K. M., Bentum, I. A., Otchere, J., Anyan, W. K., Brown, C. A., Osada, Y., Takeo, S., Kojima, S. & Ohta, N. (2004), 'Infant schistosomiasis in Ghana: a survey in an irrigation community', *Tropical Medicine & International Health* **9**(8).

Brooker, S. (2007), 'Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**(1), 1–8.

Brown, M. (2011), 'Schistosomiasis', *Clinical Medicine* **11**(5), 479–482.

Bruun, B. & Aagaard-Hansen, J. (2008), *The Social Context of Schistosomiasis and Its Control: An Introduction and Annotated Bibliography*, Special Programme for Research and Training in Tropical Diseases (TDR), World Health Organization.

Chandiwana, S. K., Woolhouse, M. E. J. & Bradley, M. (1991), 'Factors affecting the intensity of reinfection with *Schistosoma haematobium* following treatment with praziquantel', *Parasitology* **102**, 73–83.

Chitsulo, L., Engels, D., Montresor, A. & Saviola, L. (2000), 'The global status of schistosomiasis and its control', *Acta Tropica* **77 (1)**, 41–51.

Cioli, D. & Pica-Mattoccia, L. (2003), 'Praziquantel', *Parasitology Research* **90**, S3–S9.

Collett, D. (2003), *Modeling Binary Data*, Texts in Statistical Science Series, Chapman and Hall/CRC, 1ccop. 2003.

Cowper, S. (1971), *A Synopsis of African Bilharziasis*, H. K. Lewis.

Crawley, M. J. (2014), *Statistics: An introduction using R*, 2nd edition edn, Wiley-Blackwell.

Cressie, N. (1993), *Statistics for spatial data*, Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley.

de Vlas, S. & Gryseels, B. (1992), 'Underestimation of *Schistosoma mansoni* prevalences', *Parasitology Today* **8**(8), 274–277.

Diggle, P. J. & Ribeiro, P. J. (2007), *Model-based Geostatistics*, Springer.

Doehring, E., Vester, U., Ehrich, J. & Feldmeier, H. (1985), 'Circadian variation of ova excretion, proteinuria, hematuria, and leukocyturia in urinary schistosomiasis', *Kidney International* **27**(4), 667–671.

Eagle, S. & Diggle, P. J. (2012), Spatial independence test for Geo-statistical data, Technical report, University of Liverpool.

Engels, D., Chitsulo, L., Montresor, A. & Savioli, L. (2002), 'The global epidemiological situation of schistosomiasis and new approaches to control and research', *Acta Tropica* **82**(2), 139–146.

Farooq, M. & Mallah, M. (1966), 'The behavioural pattern of social and religious water-contact activities in Egypt-49 bilharziasis project area', *Bulletin of the World Health Organization* **35**, 377–387.

Fenwick, A., Webster, J. P., Bosque-Oliva, E., Blair, L., Fleming, F. M., Zhang, Y., Garba, A., Stothard, J. R., Gabrielli, A. F., Clements, A. C. A., Kabatereine, N. B., Toure, S., Dembele, R., Nyandindi, U., Mwansa, J. & Koukounari, A. (2009), 'The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002-2008', *Parasitology* **136**(13), 1719–1730.

Gelfand, A., Diggle, P., Guttorp, P. & Fuentes, M. (2010), *Handbook of Spatial Statistics*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis.

Gray, D. J., McManus, D. P., Li, Y. S., Williams, G. M., Bergquist, R. & Ross, A. G. (2010), 'Schistosomiasis elimination: lessons from the past guide the future', *Lancet Infectious Diseases* **10**(10), 733–736.

Hilbe, J. (2009), *Logistic Regression Models*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Husein, M., Talaat, M., El-Sayed, M., El-Badawi, A. & Evans, D. (1996), 'Who misses out with school-based health programmes? a study of schistosomiasis control in

Egypt', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**(4), 362 – 365.

Jordan, P. & Webbe, G. (1982), *Schistosomiasis: epidemiology, treatment, and control*, Heinemann.

Jr, P. J. R. & Diggle, P. J. (2012), *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.2.

Kahle, D. & Wickham, H. (2013), *ggmap: A package for spatial visualization with Google Maps and OpenStreetMap*. R package version 2.3.

Lammie, P. J., Fenwick, A. & Utzinger, J. (2006), 'A blueprint for success: integration of neglected tropical disease control programmes', *Trends in Parasitology* **22**(7), 313 – 321.

Lustigman, S., Prichard, R. K., Gazzinelli, A., Grant, W. N., Boatin, B. A., McCarthy, J. S. & Basanez, M. G. (2012), 'A research agenda for helminth diseases of humans: The problem of helminthiases', *Plos Neglected Tropical Diseases* **6**(4).

Montresor, A., Odermatt, P., Muth, S., Iwata, F., RajaŠa, Y., Assis, A., Zulkifli, A., Kabatereine, N., Fenwick, A., Al-Awaidy, S., Allen, H., Engels, D. & Savioli, L. (2005), 'The WHO dose pole for the administration of praziquantel is also accurate in non-African populations', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **99**(1), 78 – 81.

Mott, K. E., Desjeux, P., Moncayo, A., Ranque, P. & de Raadt, P. (1990), 'Parasitic diseases and urban development', *Bulletin of the World Health Organization* **68**(6), 691–698.

Myers, J. C. (1997), *Geostatistical error management: quantifying uncertainty for environmental sampling and mapping.*, International Thomson Publishing.

Odogwu, S. E., Ramamurthy, N. K., Kabatereine, N. B., Kazibwe, F., Tukahebwa, E., Webster, J. P., Fenwick, A. & Stothard, J. R. (2006), '*Schistosoma mansoni* in infants (aged <3 years) along the Ugandan shoreline of Lake Victoria', *Annals of Tropical Medicine and Parasitology* **100**(4), 315–326.

Okwuosa, V. & Osuala, F. (1993), 'Toxicity of washing soaps to *Schistosoma mansoni* cercariae and effects of sublethal concentrations on infectivity in mice', *Applied Parasitology* **34**(1), 69–75.

Peng, W., Tao, B., Clements, A., Jiang, Q., Zhang, Z., YB, Z. & Jiang, Q. (2010), 'Identifying high-risk areas of schistosomiasis and associated risk factors in the Poyang Lake region, China', *Parasitology* **137**, 1099–1107.

Rabello, A. (1997), 'Diagnosing schistosomiasis', *Mem Inst Oswaldo Cruz, Rio de Janeiro (Memórias do Instituto Oswaldo Cruz)* **92**(5), 669–676.

Richter, J. (2003), 'The impact of chemotherapy on morbidity due to schistosomiasis', *Acta Tropica* **86**, 161–183.

Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. (2006), 'Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk', *The Lancet Infectious Diseases* **6**(7), 411 – 425.

Stothard, J. R. & Gabrielli, A. (2007), 'Schistosomiasis in African infants and preschool children: to treat or not to treat?', *Trends in Parasitology* **23**(3), 83 – 86.

Stothard, J. R., Sousa-Figueiredo, J. C., Betson, M., Bustinduy, A. & Reinhard-Rupp, J. (2013), 'Schistosomiasis in african infants and preschool children: let them now be treated!', *Trends in Parasitology* **29**(4), 197 – 205.

Talaat, M. & Evans, D. (2000), 'The costs and coverage of a strategy to control schistosomiasis morbidity in nonenrolled school-age children in Egypt', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **94**(4), 449 – 454.

UNESCO (2012), *Global Education Digest: Comparing Education Statistics Across the World, UNESCO Reference Works 2012*, UNESCO Reference Works Series, UNESCO Publishing.

Utzinger, J., Raso, G., Brooker, S., de Savigny, D., Tanner, M., Ornbjerg, N., Singer, B. H. & N'Goran, E. K. (2009), 'Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution', *Parasitology* **136**(13), 1859–1874.

Vounatsou, P., Raso, G., Tanner, M., N'goran, E. K. & Utzinger, J. (2009), 'Bayesian geostatistical modelling for mapping schistosomiasis transmission', *Parasitology* **136**, 1695–1705.

VRA (2012), 'Non-power activities', `http://www.vra.com/`. Accessed: 12/04/2012.

Waller, L. & Gotway, C. (2004), *Applied Spatial Statistics for Public Health Data*, Vol. 368, Wiley-Interscience.

Webster, R. & Oliver, M. (2001), *Geostatistics for environmental scientists*, Statistics in practice, John Wiley & Sons.

White, P., Webber, D. J. & Helvin, A. (2014), *Modellling structural change using broken sticks*, University of the West of England, Bristol, UK.

WHO (1985), 'The Control of Schistosomiasis', `http://whqlibdoc.who.int/trs/WHO_TRS_728.pdf`. Accessed: 12/04/2012.

WHO (1987), 'Atlas of the global distribution of schistosomiasis: 15 - GHANA', `http://whqlibdoc.who.int/temp/schistosomiases/2867810604_15.pdf`.

WHO (2005), 'Report of the Scientific Working Group Meeting on Schistosomiasis: 14-16 November 2005, Geneva, Switzerland', `http://www.who.int/tdr/publications/documents/swg_schisto.pdf`.

WHO (2006*a*), *Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*, World Health Organization, Geneva, Switzerland.

WHO (2006*b*), *Rapid assessment and decision charts for schistosomiasis and soil-transmitted helminthiasis. In: Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*, World Health Organization, Geneva, Switzerland.

WHO (2009), Report of an informal working group meeting on urogenital schistosomiasis and HIV transmission, Technical report, World Health Organization, Geneva, Switzerland.

WHO (2010), 'Informal consultation on expanding schistosomischi control in Africa', http://www.who.int/schistosomiasis/epidemiology/PZQ_WHO_report_meeting.pdf.

WHO (2013*a*), 'Schistosomiasis: number of people treated in 2011', *Weekly Epidemiological Record* **88**(8), 81–88.

WHO (2013*b*), *Schistosomiasis: Progress report 2001-2011 and strategic plan 2012-2020*, World Health Organization, Geneva, Switzerland.

Wikipedia (2014), 'Akosombo Dam', http://en.wikipedia.org/wiki/Akosombo_Dam. [Online; accessed 27-Dec-2014].

Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC.

Woolhouse, M. E. J., Etard, J.-F., Dietz, K., Ndhlovu, P. D. & Chandiwana, S. K. (1998), 'Heterogeneities in schistosome transmission dynamics and control', *Parasitology* **117**, 475–482.

Zuur, A. (2012), *A Beginner's Guide to Generalized Additive Models with R*, Highland Statistics Limited.

Zuur, A., Hilbe, J. & Ieno, E. (2013), *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists*, A Beginner's Guide to, Highland Statistics.

Zuur, A., Ieno, E. & Smith, G. (2007), *Analyzing Ecological Data*, Statistics for Biology and Health, Springer New York.

Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. (2011), *Mixed Effects Models and Extensions in Ecology with R.*, Springer.

Part II

Transmission patterns of *S. haematobium* and the soil-transmitted helminth infections in low endemicity areas in the Greater Accra Region of Ghana

Abstract

The chemotherapy implementation strategy for targeting and treating the remaining reservoirs of schistosome transmissions in post-intervention areas has so far proved ineffective. Reasons for this may include the fact that transmissions are spatially heterogeneous; thereby increasing the transmission potential and the risk of resurgence. Hence, more robust strategies are required for the effective targeting and control of infections in these areas. The present paper investigates the role of selected socio-economic factors in predicting the transmission patterns of persisting schistosome and soil-transmitted helminth infections in some low endemicity areas in the Greater Accra Region of Ghana.

1 Introduction

The World Health Organisation recommends a dual control strategy for schistosomiasis in endemic settings. The initial phase of this strategy involves morbidity control in high transmission areas through the mass administration of chemotherapy whilst the final phase mainly focuses on the consolidation of control measures for the ensuing low transmission rates (Figure 53) (Utzinger et al. 2003, WHO 2002). Intervention programmes have, therefore, relied heavily on model-based risk maps to define the boundaries of high transmission areas in order to guide the initial allocation of intervention resources at large geographical scales (Brooker et al. 2009, 2002). Once these transmission limits are established, rapid assessments of infections are conducted at the district levels and used in establishing modalities for the implementation of chemotherapy (Brooker et al. 2002, WHO 2013*b*).

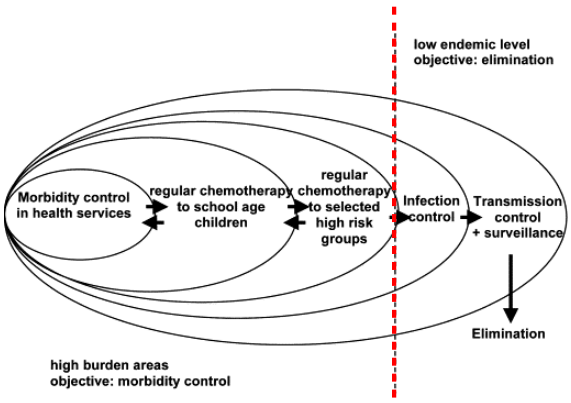


Figure 53: Stages involved in the control of schistosomiasis in high and low endemicity areas (adapted from Engels et al. 2002 with permission of the rights holder, Elsevier.)

The transition from morbidity control to the consolidation phase of control programmes begins when the prevalence of schistosome infections in post-intervention areas falls below 10%. While the administration of chemotherapy continues in these low endemicity areas, the frequency and scope of administration is considerably decreased. For instance, school-aged children are scheduled to receive only two treatments during the entire period of their primary school education. Moreover, provisions for treating symptomatic cases in the general population in these areas are restricted to health care centres (Clements et al. 2006, WHO 2013*a*, Lustigman et al. 2012).

Therefore, this strategy of targeting the remaining reservoirs of transmissions only results in the treatment of about 33% of school-aged children in post-intervention areas. Moreover, most of the infected cases in these low endemicity areas tend to harbour light-intensity infections which are often characterised by subtle morbidities with non-specific symptoms (Lustigman et al. 2012). There is, therefore, the tendency for potentially infected individuals to overlook these non-discernible symptoms and consequently fail to seek treatment. Additionally, the poor resource allocation in endemic settings, coupled with the slow pace of socio-economic development, has presented inherent difficulties for control programmes to integrate the operational components of transmission control such as improved sanitation and potable water (Gazzinelli et al. 2012, WHO 2013*b*). Therefore, the strategies in place for sustaining the low transmission rates in post-intervention areas may be inadequate in preventing the resurgence of infections.

Indeed, studies in many post-intervention areas have highlighted the high risk of resurgence as control programmes transition from morbidity control to their consolidation phases. Empirical evidence from chemotherapy campaigns in highly endemic countries such as Egypt, Brazil, China and Mali have lent credence to this fact (WHO 2005, Clements et al. 2009). Moreover, the fact that sub-Saharan Africa still harbours about 80% of the global burden of schistosomiasis is in itself proof that the chemotherapy intervention campaigns that took place across the continent in the late 1980's and 1990's were not very successful in sustaining the low transmission rates either (Lammie et al. 2006, WHO 2002).

Therefore, more robust strategies for tackling the remaining reservoirs of transmissions in post-intervention areas would be required if the ongoing chemotherapy intervention campaigns in sub-Saharan African were to achieve their targets of reducing transmission rates below the critical break point. However, while the need for adjusting the current strategy to suit the changing epidemiological conditions in post-intervention areas is well-recognised and emphasised by the World Health Organisation (WHO 2013*b*), the strategies in place for identifying these reservoirs of transmission have proved less effective. This is mainly due to the fact that the transmission of schistosomiasis is spatially heterogeneous. Moreover, the degree of this heterogeneity tends to vary within different endemic settings in response to specific micro-level factors (Brooker 2007, Bruun &

Aagaard-Hansen 2008). Therefore, these clusters of transmissions end up assuming an even more sparse distribution when endemic areas are subjected to chemotherapeutic pressure during intervention campaigns (Basanez et al. 2012). These sparse clusters of transmissions, which tend to increase the transmission potential of infections, therefore present additional challenges to control due to the increased risk of resurgence (Woolhouse et al. 1998).

The life cycle of the schistosome species is, however, such that they undergo the asexual phase of their reproduction in the environment, rather than within the human definitive host (Figure 54) (Gryseels et al. 2006). Therefore, every schistosome parasite in an infected human host is attributable to an infective event that involves direct contact with the infective free-living larval stages of the parasite in their freshwater habitats. It, therefore, follows that the level of persistence of schistosome infections within any post-intervention setting could be a direct function of the availability and level of access to basic infrastructure such as potable water and improved sanitation. Hence, the exact extent of the spatial heterogeneity in the distribution of the remaining reservoirs of transmission within any specific post-intervention setting could be directly linked to the socio-economic inequalities across households that determine access to improved sanitation and potable water.

Therefore, the primary objective of the present study are: to investigate the persistence of schistosome infections in selected post-intervention areas in Ghana as a function of the socio-economic standards of households; and to use the resulting model in the prediction of the remaining reservoirs of transmission in the study districts as well as the surrounding areas. Moreover, intervention programmes usually adopt an integrated strategy whereby multiple tropical infectious diseases that occur at the same geographical unit are targeted concurrently during the implementation of chemotherapy. Therefore, interventions for schistosomiasis and the soil-transmitted helminth infections (STH), which tend to be co-endemic and share the same susceptible human host population (Figure 55), are implemented concurrently (Brooker et al. 2009). We, therefore, extend our study objectives to cover the soil-transmitted helminth infections.

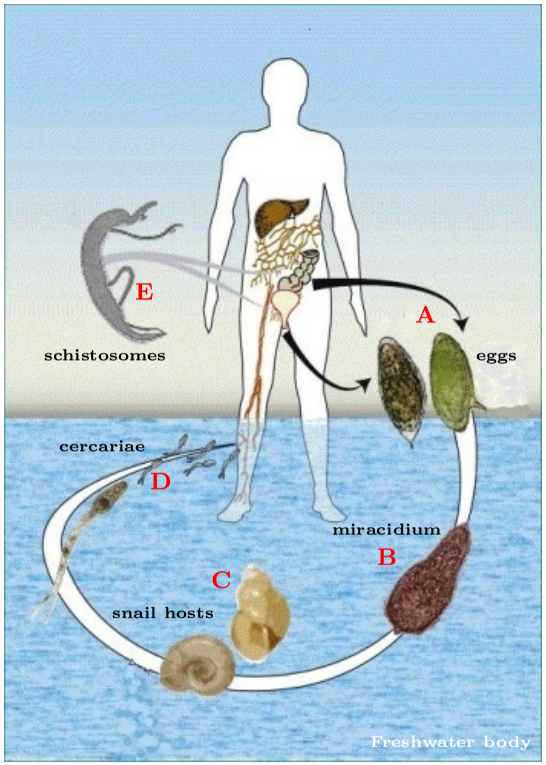


Figure 54: Transmission cycle of *Schistosoma haematobium* and *S. mansoni* (adapted from Gryseels et al. 2006). A: Schistosome eggs, excreted through the urine and faeces of infected individuals, enter freshwater bodies. B: The eggs hatch to release the first larval stages, miracidia, into the water. C: The larvae swim about until they locate and penetrate the intermediate snail host where they reproduce asexually. D: The second larval stages leave the snails and spin about in the water until they find the human definitive host. Infection occurs through active penetration of the skin. E: The larvae migrate through the body and develop into adult worms. The adult worms lay eggs which continue the cycle.

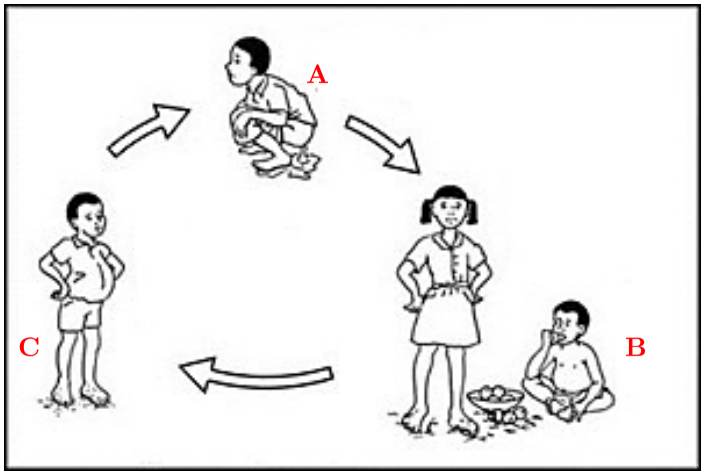


Figure 55: Schematic diagram of the transmission cycle of soil-transmitted helminthiasis (adapted from WHO 2011). A: The soil gets contaminated with STH eggs from the faeces of an infected individual. These eggs develop in the soil. B: Infection is contracted through the ingestion of the embryonated eggs in food such as raw vegetables, eating with dirty hands or active penetration of the skin by the larvae of hookworms. C: Once inside the body, the eggs/larvae develop into adult worms which produce more eggs to continue the cycle.

2 Materials and Methods

This study is, therefore, based on the hypothesis that the persistence of schistosome infections in post-intervention areas is principally governed by the socio-economic conditions of households that determine access to improved sanitation and potable water. Moreover, the soil-transmitted helminth infections tend to persist under conditions of poverty and inadequate hygiene practices that result in the environmental contamination and exposure to the infective stages of the parasites. We are, therefore, arguing that the transmission of the schistosomes and soil-transmitted helminth infections would follow similar patterns within any given endemic setting in response to the socio-economic standards of households.

2.1 Data Source

To test our study hypotheses, we made use of a pre-collected cross-sectional survey data from three selected districts of the Greater Accra Region of Ghana (GLOFAL 2006). Our choice of the dataset was influenced by two main factors. Firstly, the study population for the original project for which the data were collected comprised of children aged 4-17 years. This age-group constitutes the most susceptible human host population for both schistosomes and soil-transmitted helminth infections, and hence a suitable choice for this study. Indeed, children in this age-group have been the target for large scale chemotherapy interventions by organisations such as the Partners for Parasite Control and the Integrated Control of Schistosomiasis in sub-Saharan Africa. Secondly, the data were collected from areas which are known to have low schistosome transmission rates.

2.2 Study Sites

Our study sites, the Dangme East (DE), Accra Metropolitan Area (AM) and the Ga East (GE) districts of the Greater Accra Region (Figure 56), were respectively rural, urban and peri-urban areas with low schistosome transmission rates. The rural sites had received multiple school-based chemotherapy interventions for schistosomiasis and

the soil-transmitted helminth infections in the past. The last of such interventions was administered about three years prior to the data collection. However, the urban and peri-urban sites had only received school-based chemotherapy interventions for the soil-transmitted helminths infections. The next two sections present detailed descriptions of these sites.

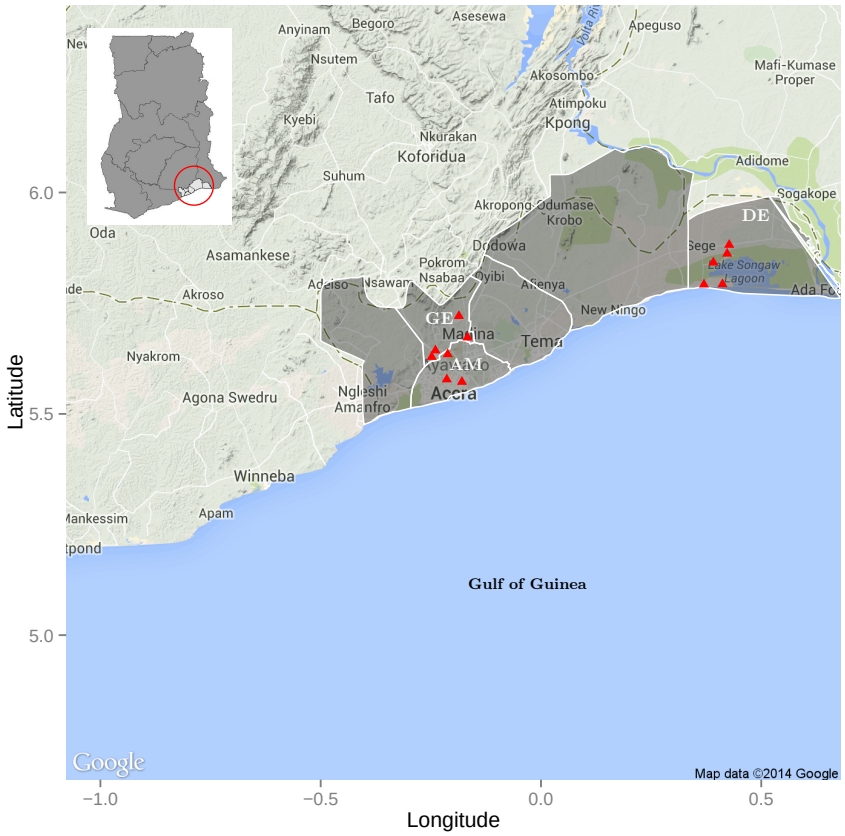


Figure 56: Main: Map of the Greater Accra Region showing the locations of the recruited schools (i.e. the red triangles). Inset: Map of Ghana showing the location of our study region (circled). Source of shapefiles: Map Library 2014. `R` Packages used in map preparation: `ggmap` (Kahle & Wickham 2013).

2.2.1 The Urban and Peri-Urban Sites

As the capital city of Ghana, Accra is on the receiving end of the majority of the migration within the country. With its annual growth rate of 3.1%, it is estimated to be home to about 16% of Ghana's population. Therefore, the total population of Accra, at the time of the 2010 population census, was around 1.8 million (CRED 2013, GSS 2012). The resulting pressure imposed by this influx of people on social amenities and land use has respectively led to the rationing of potable water throughout the city and the development of illegal structures, including slums, in waterways.

The most affected river, as far as building in waterways is concerned, is the Odaw River which drains most of the central parts of Accra (Figure 57). The development of slums on the course of this river, together with its pollution with sewage and solid waste from these slums, affects the drainage of the river. Therefore, flooding of the catchment areas of this river is common during the rainy season (CRED 2013). The diverse ecological niches of the intermediate snail hosts of the schistosome species are, however, known to include stagnant waters with high levels of organic pollution (Mott et al. 1990, Cowper 1971), thereby making the Odaw River a potential habitat for these snails.

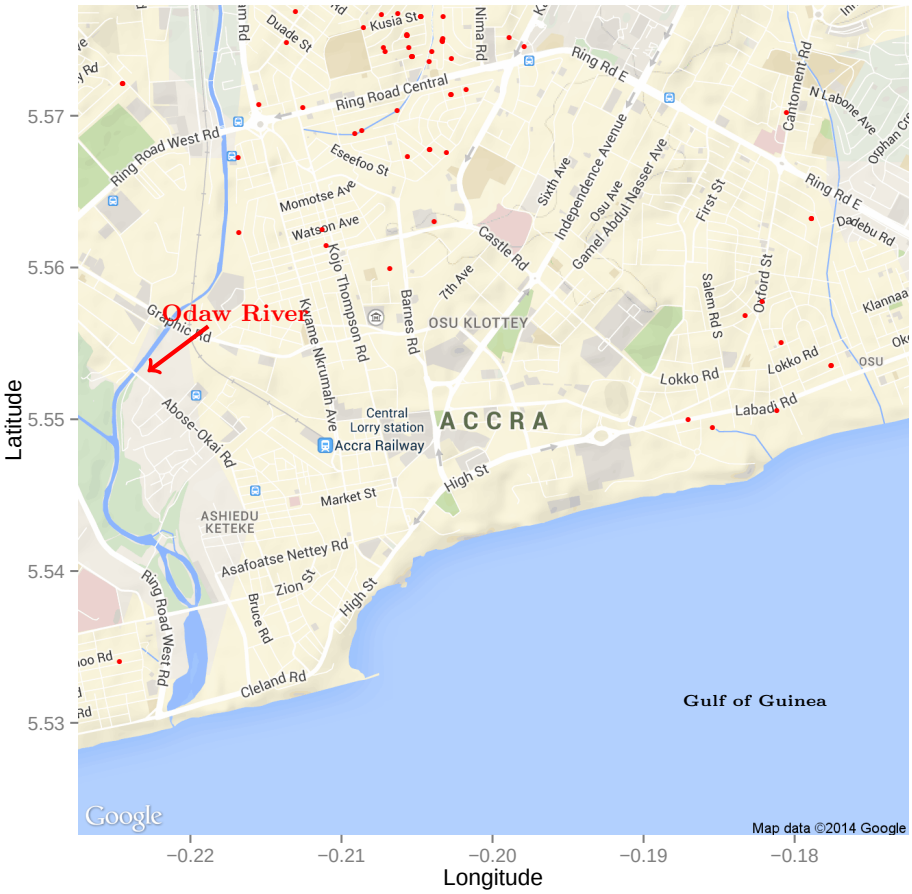


Figure 57: A section of the urban site showing the Odaw River and the locations of the homes of some of our study participants (i.e. the red dots). This River also flows through parts of the peri-urban site.

Therefore, as far as the transmission of schistosome infections is concerned, the following scenarios may apply to the urban and peri-urban sites in this study: infected migrants from rural areas may move into areas in close proximity to the Odaw River, which harbours the snail vectors, and initiate new transmission foci; infected migrants may move into other parts of the city where these snail vectors are non-existent; the flooding

of the Odaw River during the rainy season may increase the transmission potential by extending the geographical limits of infections within these areas; and lastly, non-infected children from these urban and peri-urban areas may spend their school holidays in other endemic parts of the country and return with acute infections.

However, aside from the pollution by human excreta, industrial wastes were also discharged into the river in its lower catchment areas. Unlike the organic wastes, however, these industrial discharge are known to have a deleterious effect on the intermediate snail hosts species (Cowper 1971). Therefore, the effect of these industrial wastes, together with the salinity of the lagoon around the mouth of the river, may ultimately impact negatively on transmission (Mahmoud 2001).

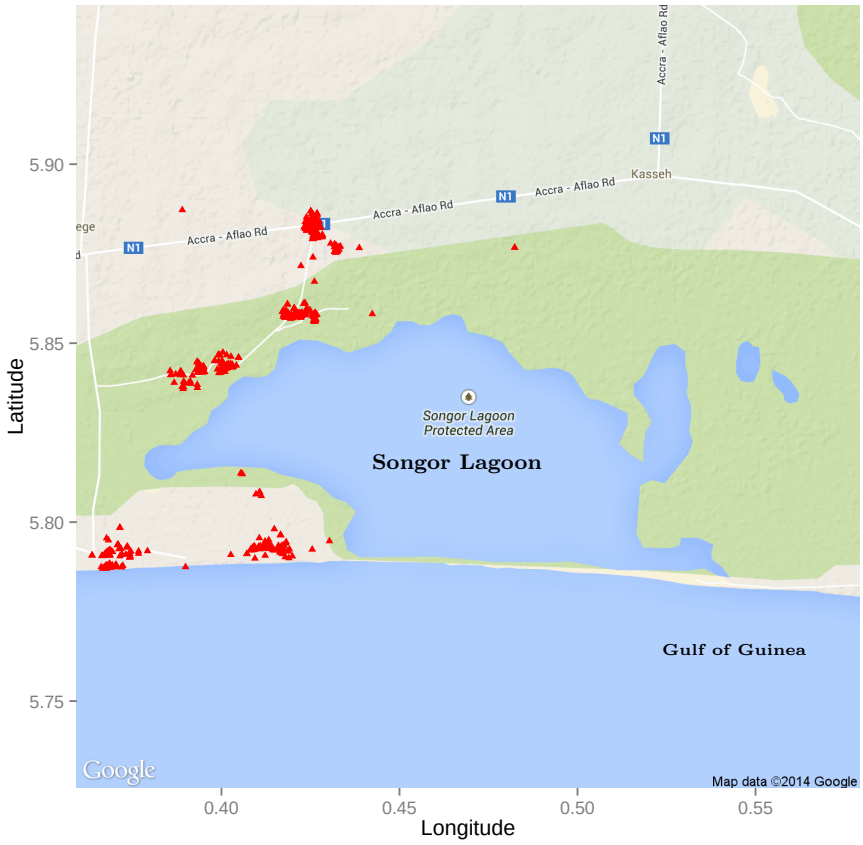


Figure 58: The locations of our five rural sites relative to the Songor Lagoon in the Dangme East District. The red triangles indicate the locations of the homes of the children in the surveyed schools.

2.2.2 The Rural Sites

Economic activities in our rural sites mainly revolved around the Songor Lagoon (Figure 58). While extended exposure to high salinity levels are generally thought to be

detrimental to the snail hosts, there are also reported accounts of the transmission of *S. haematobium* occurring in certain lagoons in West Africa (Mahmoud 2001, Cowper 1971). This therefore implies *Bulinus* spp may tolerate some levels of salinity. On the other hand, *Biomphalaria* spp rarely occurs in brackish water conditions (Mahmoud 2001). Hence, depending on its salinity levels, the Songor Lagoon may or may not have been a source of *S. haematobium* transmission.

However, economic activities associated with the Lagoon such as salt extraction, the harvesting of tilapia and vegetable farming may have had a direct influence on the socio-economic standards of families in these communities (Wikipedia 2014*b*). Given that these communities lacked social amenities such as potable water supply, the use of alternative water sources could not have really reflected the socio-economic standards of the different households.

2.3 Recruitment Strategy

In this section, we examine the recruitment strategy employed by the original study for which the data were collected. This is to enable us to assess how the recruitment strategy may influence our data analysis strategy as well as the interpretation of our findings.

Prior to recruiting the schools for the main GLOFAL project, pilot surveys were conducted in selected parts of the Greater Accra Region. The following sections outline the details of the pilot surveys and the recruitment criteria.

2.3.1 Rural Site

Schools in the rural parts of the Dangme East District with post-intervention histories were targeted for the pilot survey. The recruited schools were chosen using the criteria as outlined in Table 27.

Table 27: Criteria for recruiting schools in the rural sites

Number of schools targeted:	13 schools
School selection criterion:	History of school-based chemotherapy intervention in the 3 years prior to the study.
Sampling strategy:	Systematic sampling of every other child in the third-year class.
Assessment of infections:	Urine samples from the selected children were analysed by standard parasitology techniques for the eggs of *S. haematobium*.
Recruitment criterion for schools:	*S. haematobium* infection prevalence of $\leq 10\%$

2.3.2 Urban and Peri-Urban Sites

There were time limits between the field collection and laboratory analysis for some of the samples that were collected during the original project. Therefore, the selection of schools in the urban and peri-urban areas was mainly based on their proximity in time i.e. the ease with which samples could be transported to the Noguchi Memorial Institute for Medical Research, where the laboratory analysis of the samples occurred. We do not, however, regard this proximity in time as a potential source of selection bias.

Recorded infection levels during a previous study (Hogewoning et al. 2010) in the peri-urban area as well as the willingness of the schools to participate served as additional criteria for recruitment. The number of recruited schools in each of the study sites are presented in Table 28 below.

2.3.3 Sample Population

All enrolled children in primary and junior secondary in the recruited schools were initially considered for inclusion in the main GLOFAL study. This inclusion criterion was, however, modified to an age range of 4-17 years when children in the rural site were found to be significantly older. The selection criteria were restricted to written informed consents from the parents or guardians. Moreover, the school recruitment process in the urban site was largely influenced by aspects of the allergy component of the main GLOFAL project that involved the collection of blood samples from the participants. Table 28 presents the number of recruited children.

Table 28: Recruited schools in the various study sites and the overall response rate. The "Enrolment" column only refers to the number of enrolled children who fell within the age range for the study.

School / Site	School Status	Enrolment	Number Recruited	Response Rate (%)
Rural				
Koluedor	State School	300	173	
Toflokpo	State School	380	136	
Agbedrafor	State School	200	160	
Anyamam	State School	300	291	
Goi	State School	200	122	
Total		**1380**	**882**	63.91
Peri - Urban				
Pantang	State School	352	168	
Emmanuel Presbyterian School	State School	650	250	
Nii Okine Basic School	State School	300	130	
Total		**1302**	**548**	42.09
Urban				
Greenhill International School	Private School	900	287	
Mona Lisa School	Private School	180	43	
De Youngster's International	Private School	740	203	
Morning Star International School	Private School	800	47	
Total		**2620**	**580**	22.14

2.4 Ethical Considerations

The original study was reviewed and approved by the Institutional Review Board of the Noguchi Memorial Institute for Medical Research in Accra, Ghana. Signed informed consents by parents or guardians were used as the main criterion for including children in the study. Formal approvals for recruiting state schools in the rural and peri-urban areas were granted by the Directors of Education for the Dangme East and Ga East Districts of the Greater Accra Region.

2.5 Measured Exposures

Going back to our study hypothesis, we are arguing that the persistence of schistosome infections in post-intervention areas may be a direct function of the socio-economic standards of households which subsequently determine access to improved sanitation and potable water. Therefore, in a bid to capture some of the factors that reflect the socio-economic inequalities across households in our study sites, we considered three main categories of exposure variables: individual-level covariates; some of the commonly used indicators of the socio-economic standards of households in low income countries (Huang & Manderson 1992, Rutstein & Johnson 2004); and health-seeking behaviours

at the household level (Table 29). Our questionnaire on migratory patterns did not capture this information as well as we had hoped. Therefore, information on migration was excluded from the analysis.

Table 29: Main categories of measured exposures (collected at baseline in an interviewer-administered questionnaire* survey that targeted parents/ guardians of the recruited children).

Individual-Level Covariates	Indices of Socio-Economic Standards	Periodic Deworming
Age	Source of water	Period since last anthelmintic treatment
Sex	Type of sanitation	
	Type of house	
	Building material	
	Source of fuel for cooking	
	Electricity supply in the home	
	Means of transportation to school	
	Educational level of family provider	

* A sample of the questionnaire is attached as Appendix C

But how well can these factors be used in assessing the socio-economic standards of households in our study sites? In the next section, we examine some of the indices that have been developed specifically for assessing the socio-economic standards of households in developing countries. We will then determine which of these indices would best fit into the context of this study.

2.5.1 Indices of Socio-Economic Standards in Low Income Countries

Three main indices, namely income levels, consumption expenditure and wealth are widely used in assessing the socio-economic standards of households. However, the inherent difficulties associated with the accurate measurement of income levels and the consumption expenditure of households in low income countries has made wealth the preferred index of assessment (Rutstein & Johnson 2004).

Therefore, the relative wealth of households is commonly assessed with variables that are generally regarded as indicators of their permanent economic conditions. These indicator variables, which may include assets ownership (radio, television, refrigerator,

type of vehicle, etc.), housing characteristics (type of roofing, wall material, et cetera.) and access to utility services are assumed to be correlated with the relative positions of households on an underlying wealth scale (Figure 59).

Since the wealth index is purely asset based, other commonly used measures of socio-economic standards such as occupation and educational level are excluded from the assessment. Concerns have, however, been raised in some circles, about the ability of this wealth index to effectively capture the economic standards of both rural and urban households. It has been argued that most of the indicator variables in this wealth index may be biased against rural households where asset ownership may be influenced by the availability of public utility services (Rutstein 2008).

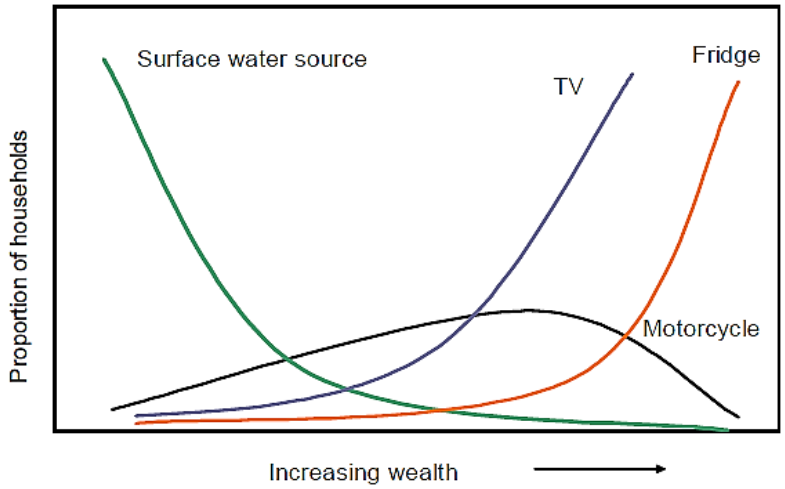


Figure 59: The assumed distribution of household asset ownership with increasing wealth in low income countries (Adapted from (Rutstein & Johnson 2004)). The relative position of households on this wealth scale determines their access to improved services.

Moreover, the inclusion of utility services in the wealth index is based on the assumption that wealthy families tend to acquire properties in areas where such services are provided (Rutstein & Johnson 2004). However, this assumption may not always hold for many households in the peri-urban areas of developing countries where families acquire land and develop their own properties. In such cases, the availability of land, rather than the provision of public utility services in the area, is usually used as basis for the selection of potential dwellings by families.

Therefore, the provision of improved services such as potable water and sanitation facilities may not always correlate with the position of households on the wealth index. Regardless of these issues, however, most of the indicator variables for wealth could directly be linked to the health outcomes of households in endemic areas, thereby making it a suitable index for health research studies such as ours.

2.5.2 Household Asset-Based Indices

For this method, the first principal components of the indicator variables for wealth are most often used as the underlying wealth scale and the relative position of any household on this scale is determined by computing the weights of the principal components (PCA) (Rutstein & Johnson 2004).

Apart from the PCA, other methods of establishing the economic status of households have included the inverse proportion approach which assigns the highest weight to the least owned asset. The main assumption behind this method is that the most expensive assets can only be acquired by wealthier households and hence would generally be uncommon in many other households. Finally, there is the hierarchical ordered probit strategy which assigns weights to indicator variables on the basis of their position on an assumed unmeasured underlying wealth scale (Rutstein & Johnson 2004).

2.5.3 The Chosen Index for this Study

Rather than building a household asset index, we chose to enter all the measured exposures, Tables 30 and 32, into our statistical models and regard the regression parameter estimates for these variables as their weight. This "parameter-based approach" is also generally regarded as a plausible alternative method of constructing the wealth indices for households (Filmer & Pritchett 2001).

We, however, note that although this adopted strategy may have enabled us to determine the weights of the measured indicator variables for socio-economic standards, these variables may also exerted their own effects on our study outcomes. Hence, the estimated weights could not easily be interpreted as the relative wealth of households in this study.

Table 30: How the factor levels of the main categorical explanatory variables were derived

Variable	Factor Levels of variable	List of Questionnaire Responses
Source of water (see Table 31)		
	Tap	"Pipe", "rain water"
	Tanker	"Water from tanker trucks", "bottled mineral water", "sachet water"
	Well	"Water from well"
	Surface	"Dam", "pond", "river"
Type of sanitation (see Table 31)		
	Private water closet	"Private wc"
	Compound latrine	"Pit latrine", "outdoor latrine", "outdoor wc", "compound wc"
	Public latrine	"KVIP[a]", "public wc[b]", "public latrine", "chamber pot"
	Open sanitation	"Free range", "the bush", "beach", "inside dug hole"
Type of house		
	Detached	"Detached house", "mission house", "detached wooden structure", "kiosk"
	Semi-detached	"Semi-detached house", "flat", "quarters"
	Compound House	"Compound house", "family house"
Building material		
	Cement	"Cement", "brick"
	Mud	"Mud", "wood", "palm frond", "aluminium sheet"
Source of cooking fuel		
	Charcoal	"Charcoal", "kerosene"
	Firewood	"Firewood", "wood"
	Liquefied petroleum gas (LPG)	"LPG", "electricity"
Electricity supply in the home		
	No	"False"
	Yes	"True"
Means of transportation to school		
	Walk	"Walk", "bicycle"
	Bus	"Bus", "trotro[c]", "motorbike"
	Private	"Pivate car", "taxi"
Educational level of family provider		
	No formal education	"No formal education", "makaranta[d]"
	Basic	"Primary", "middle school", "junior secondary"
	Secondary	"Senior secondary", "O-level", "A-level"
	Further	"Vocational school", "training college"
	Higher	"Polytechnic", "university"

[a] KVIP: Kumasi Ventilated-Improved Pit i.e. a type of sanitation facility in Ghana.

[b] wc: water closet.

[c] Trotro: A share taxi mode of transportation in Ghana that involves 15-seater minibuses (Wikipedia 2014*a*).

[d] An Islam educational system that gives lessons in Arabic and koranic studies (JHR n.d.).

Table 31: Basis for using the chosen categories for sanitation facility and source of water in Table 30.

Variable	Definition	Considerations	
		Unimproved	**Improved**
Sanitation	"The lowest-cost option for securing sustainable access to safe, hygienic, and convenient facilities and services for excreta and sullage disposal that provide privacy and dignity, while at the same time ensuring a clean and healthful living environment both at home and in the neighbourhood of users" (UNICEF 2006).	1. Use of public sanitation facilities, where fixed charges had to be paid per visit 2. Resorting to open sanitation. 3. Compound: when shared by different families living on the same compound	1. Private: when used by members of the same family
Source of Water	Water sources are rated as improved or unimproved on the basis of the quality of drinking water they provide (UNICEF 2006).	Surface water, bottled water and water from tanker trucks (UNICEF 2006).	Piped water, borehole, protected springs and rain water (UNICEF 2006).

Table 32: Types of over-the-counter anthelmintics used by our study participants.

Name of Anthelminthic	List of Questionnaire Responses
Albendazole / Mebendazole	Albendazol suspension, Albendazole, Albendazole and ivermectin, Menthozole syrup, Zeben Albendazol, Mebendazole, Mebendazole 500mg, Mebendazole 500g tablets
Wormplex	Wormplex 400 syrup, wormplex 400, Wormplex 200, Wormplex 400 suspension
Vermox/Zentel	Vermox, Vermox syrup, Vermox and wormplex, Vermox or Zentel, Vermox or wormplex, Zentel
Other	Ketrax, Levamisole hydrochloride syrup, Levamisole hydrochloride tablet, Tanzel Milk of magnesia, Tanzon, Herb, Herbal drugs, Mission syrup, Albin, Letamox

2.6 Study Outcomes

A positive infection was regarded as the presence of one or more *S. haematobium*, *S. mansoni* or STH eggs on the urine or stool slides using the methods described in Table 34. The outcome variables that were considered in this study are presented in Table 33. *Schistosoma haematobium* and *S. mansoni* infections were considered separately for the following reasons. The two infections tend to exhibit geographical variations in their distribution with *S. mansoni* being the least occurring of the two in Ghana (WHO 1987).

Table 33: The three main study outcomes that were initially considered in this study.

Outcome Variables	Definition
S. haematobium mono-infections	Infection with only *S. haematobium*
S. mansoni mono-infections	Infection with only *S. mansoni*
STH infections	Infection with any of the soil-transmitted helminths.

Moreover, it has been suggested that the efficacy of praziquantel may differ for the two schistosome infections (WHO 2013*b*). However, setting-specific variations have also been reported in the susceptibility of these infections to praziquantel. For instance, higher cure rates have been reported for *S. mansoni* infections in Cameroon whilst *S. haematobium* seems to be the more susceptible species to chemotherapy interventions in Niger (Tchuem Tchuente et al. 2013).

Table 34: Parasitological techniques and detection methods used in processing the collected samples

Type of Infection	Sample Type	Processing Method	Detection of Infections
S. haematobium	10ml of urine	Urine filtration technique (WHO 1985)	Microscopy examination of urine slides
S. mansoni	Stool	Kato-Katz (WHO 1985)	Microscopy examination of thick stool smear slides
Soil-transmitted helminths	Stool	Kato-Katz (WHO 1985)	Microscopy examination of thick stool smear slides for hookworms (*Necator* and *Ancylostoma* spp), roundworms (*Ascaris lumbricoides*) and whipworms (*Trichuris sp*).

2.7 Exploratory Analysis

2.7.1 Criteria for Inclusion in the Analysis

Participants were considered for inclusion in the analysis if they had submitted at least one parasitological sample and taken part in the questionnaire survey. Since this analysis approach included all participants who had submitted either urine or stool samples, we resorted to fitting different models for the various outcomes, Table 33, as a way of accounting for the different sample sizes. A total of 1339 participants were considered in the analysis (Table 35).

Table 35: Number of parasitology samples collected during the school-based cross-sectional surveys in the three study sites.

	Number of Participants (%)			
Study Site	**With both Urine and Stool Samples**	**With Urine Samples Alone**	**With Stool Samples Alone**	**Total**
Rural	481 (41.04)	69 (53.08)	29 (78.38)	579
Peri-Urban	476 (40.61)	50 (38.46)	6 (16.21)	532
Urban	215 (18.34)	11 (8.46)	2 (5.41)	228
Total	**1172**	**130**	**37**	**1339**

2.7.2 Assessing Non-Linearity in Age

Non-parametric smoothers, equation 1, were employed in assessing the relationship between age and our study outcomes (Wood 2006). These estimated effects therefore formed the basis for the approximations on the forms of dependence on age in our formal models.

$$\begin{aligned} \log\left(\frac{p_i}{1-p_i}\right) &= f\left(\mathrm{X}_i\right) \\ Y_i &\sim \text{Bernoulli}(1, p_i) \end{aligned} \tag{1}$$

where p_i represents the probability that the observed outcome in the i^{th} child was positive whilst $1-p_i$ denotes otherwise; X_i is the continuous-valued covariate, age; and $f(\mathrm{X}_i)$ is the smoothing function for age in the Generalised Additive Model.

2.8 Confirmatory Analysis

2.8.1 Effect of Group Level Homogeneity

Judging by the fact that our study participants were recruited from schools, it seemed logical to examine the effects of exposures specific to these schools. This is because the home and school environments may present different exposures, hence risk behaviours may vary considerably at the school and household levels. In the following sections, we examine some of these school and household-specific exposures and their potential effect on our data analysis strategy.

2.8.1.1 Exposures Within Schools

Though health education is included in the curricula for schools, only a few schools are able to complement this health education with the necessary operational components of transmission control such as clean water supply, improved or well-maintained sanitation facilities and soap for hand washing. Hence, the level of environmental contamination and exposure to the infective stages of parasitic helminths may vary considerably for children in urban private schools and those in rural state-funded schools where access to improved sanitation and clean water may be woefully inadequate. Moreover, the perennial interruption in the supply of potable water means that the exposure levels in schools may vary depending on the availability of water storage facilities (Figure 60).

2.8.1.2 Exposures Within Households

At the household level, risk behaviours could be influenced by socio-economic conditions, and consequently the type of facilities available to household members. For instance, as part of their domestic chores, children in rural settings are normally required to collect water for the household. The risk of *Schistosoma* infection for these children may, therefore, increase when water has to be collected from surface water sources.

Moreover, a household's level of health awareness may be the direct influence of their ownership of assets such as radios and televisions that enable them to listen to health education programmes (Kloos 1995). Therefore, health-seeking habits such as regu-

lar treatment with over-the-counter anthelmintics may largely depend on knowledge and education as well as the relative position of households on the wealth index (Figure 59).

2.8.1.3 Considering Schools and Households as Random Effect Terms

Since spatial proximity is well known to influence group homogeneity (Kreft & de Leeuw 1998), the effects of the similar exposures within schools and households as well as the implications of these dependency could not be overlooked. The fundamental assumption of independence of observations in classical statistical models is violated by group-level homogeneity (Kreft & de Leeuw 1998). Hence as a way of adjusting for this dependency, we considered the option of treating households and schools as random effect terms.

(a) (b)

(c) (d)

Figure 60: Sanitation facility (a) and water storage containers (b – d) in Ghanaian state schools.

2.8.1.4 Do Schools and Households Qualify as Random Effects Terms?

Even though schools and households could logically be regarded as random effect terms (Crawley 2002), we still needed to establish if that was indeed the case for this study. Therefore, as an initial assessment to justify the specification of schools and households as random effect terms in the models, two factors were taken into consideration. Firstly, the recruited schools had to be representative of all schools that fitted into the recruitment criteria in order to qualify as random effect terms; and secondly, the clustering of participants within households needed to be sufficiently high in order for the effect of the group-level homogeneity to be significant.

However, school was ruled out as a random effect term due to the non-random school recruitment process, section 2.3. An assessment of the number of participants per household indicated that 184 houses had multiple participants ranging from 2-4 children per household (Table 36). But are these numbers sufficiently high to induce dependency in the data? To answer this questions, we performed a variance components analysis which is discussed in the next section.

Table 36: Distribution of the study participants by households in the three sites. In all, 184 households had multiple participants ranging from 2-4.

	Number of Houses			
Participants Per House	**Urban (%)**	**Peri-Urban (%)**	**Rural (%)**	**Total**
1	181 (89.16)	370 (83.90)	374 (80.43)	925
2	19 (9.36)	53 (12.02)	72 (15.48)	144
3	3 (1.48)	16 (3.63)	15 (3.23)	34
4	0 (0.00)	2 (0.45)	4 (0.86)	6
Total	**203**	**441**	**465**	**1109**

2.8.1.5 Variance Components Analysis

The contribution of the random variation due to clustering within households on the overall variance in the risk of our study outcomes was assessed as follows. Two models, a random intercept model that took the random variation into account and a fixed

effects model that completely disregarded the effect of the random variation, were fitted. A generalised likelihood ratio test was then used in assessing the comparative fit of the two models. Hence, the results of this test served as the basis for selecting the optimal models for our study outcomes. The full details of these analyses are outlined below.

2.8.1.5.1 Mixed Effects Model

This model took the random variation due to clustering within households into account. The model was, therefore, formulated as follows: the effect of factors that influenced the observed outcome in any given participant, say the jth child living in the ith household, was modelled by assuming there is as an unknown probability, p_{ij}, of acquiring an infection. Therefore, higher values of p_{ij} would signify higher chances of being infected and vice versa. Moreover, if the infection status of the jth child were denoted by a random variable, $Y_{ij} \in \{0, 1\}$, then $P(Y_{ij} = 1) = p_{ij}$ would be the probability that an infection is observed whilst $P(Y_{ij} = 0) = 1 - p_{ij}$ would denote otherwise. Hence, Y_{ij} follows the Bernoulli probability distribution (equation 2) (Collett 2003).

$$P(Y_{ij} = y_{ij}) = p^{y_{ij}}(1-p)^{1-y_{ij}}; \quad y = 0, 1 \tag{2}$$

Following the Bernoulli convention, the expected mean and variance of Y_{ij} are respectively given by:

$$E(Y_{ij}) = p_{ij} \qquad \text{and} \qquad var(Y_{ij}) = p_{ij}(1 - p_{ij})$$

To account for the random variation, U_i, the predictor function, η_{ij}, was specified as a function of both the measured covariates and U_i (equation 3). In the final step of the model formulation, the logit link function was employed in defining the relationship between pij and η_{ij} (equation 4).

$$\eta_{ij} = \beta \mathbf{X}_{ij} + \quad U_i \tag{3}$$

where $$U_i \sim N(0, \sigma^2)$$

$$p_{ij} = \frac{e^{\beta \mathbf{X}_{ij} + \quad U_i}}{1 + \quad e^{\beta \mathbf{X}_{ij} + \quad U_i}} \tag{4}$$

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) \quad = \quad \beta \mathbf{X}_{ij} \quad + \quad U_i \tag{5}$$

where β is the vector of unknown regression parameters, $\mathbf{X}_{ij}$ is the vector of measured covariates and U_i is the random variation within households.

The model was executed with the `glmer` function, as implemented with the `R lme4` package (Bates et al. 2012).

2.8.1.5.2 Fixed Effects Model

For this model, we assumed a negligible effect for the random variation, U_i. Since the variance of the random variation, σ^2, would be small under such circumstances, all households would end up having similar logistic curves (Zuur et al. 2011). Therefore, equation 5 would naturally break down into a classical fixed effects model, equation 6. Hence for this model, the unknown regression parameters were estimated by fitting a binary logistic regression model to the data.

$$\begin{gathered} \log\left(\frac{p_{ij}}{1-p_{ij}}\right) \quad = \quad \beta \mathbf{X}_{ij} \\ Y_{ij} \sim \text{Bernoulli}(1, p_{ij}) \end{gathered} \tag{6}$$

2.8.1.5.3 Assessing the Comparative Fit of Models 5 and 6

The effect of the random variation within households, U_i, on the maximum likelihood of the model formed the basis for the comparison of the two non-nested models above i.e. equations 5 and 6. Hence, a generalised likelihood ratio test that had a null and alternative hypotheses given by equation 7 was used in assessing the effect of the U_i.

$$\mathrm{H_0 : X_{ij}}\beta \qquad \text{and} \qquad \mathrm{H_A : X_{ij}}\beta + U_i \tag{7}$$

$$2 \times \log(\mathrm{L_2/L_1}) = 2\left[\log(\mathrm{L_2}) - \log(\mathrm{L_1})\right] \tag{8}$$

Therefore, given the log-likelihoods (L_1 and L_2) and number of parameters (k_1 and k_2)

in the two fitted models (equations 5 and 6); the change in deviance (equation 8) given that the null hypothesis is true would have a Chi-square (χ^2) distribution with k_1 - k_2 degrees of freedom (Pinheiro & Bates 2000, Leyland & Goldstein 2001).

A statistically significant p-value, on $\chi^2_{k_1-k_2}$ distribution, was therefore interpreted as meaning U_i gave a significant improvement in the maximum likelihood of the model, in which case the mixed effects model would be the preferred choice. Though this method of assessing the significance of the random variation only holds for large samples (Pinheiro & Bates 2000), it is generally regarded as being too conservative. This problem is, however, resolved by multiplying the nominal p-value by 0.05 (Self & Liang 1987).

2.8.1.6 Modelling the Risk of *S. haematobium* and the STH Infections

The results of the generalised likelihood ratio test, as described in section 2.8.1.5 above, provided compelling evidence ($p < 0.001$) in favour of the alternative hypothesis for *S. haematobium* mono-infections, $\mathrm{H_A} : \mathrm{X_{ij}}\beta + U_i$. Therefore, the mixed effects model was deemed suitable for the *S. haematobium* mono-infections. In the case of the STH infections, however, there was no compelling evidence ($p = 0.43$) for rejecting the null hypothesis, $\mathrm{H_0} : \mathrm{X_{ij}}\beta$. Hence, the classical fixed effects model was considered optimal for modelling the risk of the STH infections. The following sections, therefore, outline the stages of our formal regression analyses. We, however, defer the discussion on our model for *S. haematobium* mono-infections till later in section 3.2.3.

2.8.1.6.1 Model Selection

Simple logistic regression analyses were initially performed to assess the worth of each measured covariate. These covariates were then entered into the model in the order of their contribution to the overall deviance, as judged by a deviance-based test. The relative effects of the terms in the multiple logistic regression model were also assessed by a deviance-based test (Hilbe 2009). Therefore, terms that were significant on entry into the model but whose effects became non-significant after adjusting for the effects of other terms were considered for exclusion. The exclusions of any such terms were, however, effected only when their omissions did not significantly in-

crease the overall deviance of the model. The resulting model, after the backwards selection of terms, became our main effects model for the risk of the STH infections.

The terms in the main effects model were used in investigating the effect of two-way interactions. The effects of these interaction terms were assessed by computing the deferences in deviance between the nested model, i.e. the main effects model, and each of the full models containing the interaction terms, in addition to the terms in the main effects model. Non-significance interaction terms were dropped if their exclusions did not significantly increase the overall deviance of the model. The resulting model became our provisional model for the risk of the STH infections.

2.8.2 Goodness-of-Fit Check

2.8.2.1 Assessment of Over-dispersion

Following recommendations by Zuur et al. 2013 that over-dispersion in Bernoulli GLM be defined by the Pearson dispersion statistic, equation 9 was used in assessing our provisional models for extra-dispersion. Our cut-off limit for over-dispersion was set at 1 (Zuur et al. 2013).

$$\frac{\sum(y_i - \hat{p_i})}{N - k} \tag{9}$$

where y_i is the observed value of the outcome in the i^{th} child; $\hat{p_i}$ is the probability that infection was observed in the i^{th} child; N denotes the number of observations in the model and k is the number of parameters in the model, including the intercept.

2.8.3 Model Validation

2.8.3.1 Assessing the Homoscedasticity Assumption

In assessing our provisional models for non-constance error variance, the Pearson residuals, equation 10, were first sorted into groups of 10 based on the order of their sorted fitted values. The mean of the Pearson residuals in each group was then computed and plotted against the fitted values.

$$\epsilon_i = \frac{Y_i - \hat{p}_i}{\sqrt{\hat{p}_i\,(1-\hat{p}_i)}} \tag{10}$$

where Y_i is the observed response for the ith participant, $\hat{p}_i$ is the fitted value corresponding to Y_i and $\sqrt{\hat{p}_i\,(1-\hat{p}_i)}$ is an estimate of the standard deviation of the raw residuals.

2.8.3.2 Assessing the Assumption of Independence

Building on our study assumptions, we expect the transmission of schistosome and the soil-transmitted helminth infections in our low endemicity sites to be influenced by the socio-economic inequalities across households. It is, therefore, logical to expect these infections to occur in patches, provided households with similar socio-economic standards were situated in close proximities to each other. Therefore, the residuals, which represent the variation in the outcomes that are attributable to unmeasured factors and the random noise in the data (Arlinghaus 1996), would also be expected to exhibit spatial heterogeneity. In the following sections, we employ classical geostatistical concepts in investigating and describing the spatial dependence due to the heterogeneous patterns of transmission and in predicting transmissions at unsampled locations in the study region.

2.8.4 Classical Geostatistical Concepts

We, therefore, proceeded to predict the reservoirs of schistosome and STH transmissions across the three post-intervention districts under consideration (Figure 56), using information from the sampled locations, $\mathbf{s}$. Following classical geostatistical convention, we began by regarding the spatially-referenced residuals of our provisional models as single realisations, $\mathbf{z}(\mathbf{s}_i)$, of an underlying spatially continuous random process, $\mathbf{Z}(\mathbf{s}_i)$, that is specified by equation 11 (Zuur et al. 2007, Waller & Gotway 2004).

$$E\left[\mathbf{Z}(\mathbf{s})\right] = \mu \tag{11}$$

$$Cov\Big(\mathbf{Z}(\mathbf{s}_j), \mathbf{Z}(\mathbf{s}_k)\Big) = C(\mathbf{s}_j - \mathbf{s}_k)$$

where $C(.)$, the covariance function, measures the spatial autocorrelation between sampled locations $\mathbf{s}_j$ and $\mathbf{s}_k$. If the difference between sampled locations were assumed to be second-order stationary, then the mean, μ, is independent of location and the covariance/variogram only depends on the separation distance between $\mathbf{s}_j$ and $\mathbf{s}_k$ (Waller & Gotway 2004).

2.8.4.1 Structural Analysis

The spatial autocorrelation between sampled locations in the study region was quantified by computing empirical variogram for the Pearson residuals of the provisional models (equation 12). Details of the computation of the variogram has been discussed in section 2.6.3.2.5 in Part I of this thesis.

$$\gamma(\mathbf{h}) = \frac{1}{2p(\mathbf{h})} \sum_{\alpha=1}^{p(\mathbf{h})} \{z_i(\mathbf{s}_j) - z_j(\mathbf{s}_k)\}^2 \tag{12}$$

where $p(\mathbf{h})$ denotes the numbers of pairs of Pearson residuals separated by $\mathbf{h}$, whilst spatial lag, $\mathbf{h}$, is the separation distance between any given set of pairs of residuals; and z_i and z_j are the residuals for the ith and jth participants at locations $\mathbf{s}_j$ and $\mathbf{s}_k$, respectively.

2.8.4.2 Estimating the Covariance Parameters

The computed empirical variogram in section 2.8.4.1 is, however, based on lag distances between the sampled locations. Therefore, to extend these computations to the entire study region, including the non-sampled areas, a theoretical variogram model was fitted to the empirical variogram. The fitted theoretical variogram model was then used in estimating initial values of the parameters of the covariance function.

In finding the appropriate theoretical variogram model, covariance functions from the Matérn family, equation 13, were fitted onto the empirical variogram. Different values of the shape parameter, κ, were sequentially employed in fitting the covariance function until the shape of the isotropic empirical variogram was adequately captured (Figure 61). The elements of the covariance structure, Figure 62, were estimated by weighted least

squares using specific functions in the `R geoR` package (Ribeiro Jr & Diggle 2001, Diggle & Ribeiro 2007).

$$\rho(u) = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi) \tag{13}$$

where $K_{\kappa}(.)$ represents a modified Bessel function of order κ; the order, κ, determines the smoothness of the underlying spatial process; φ is a scale parameter.

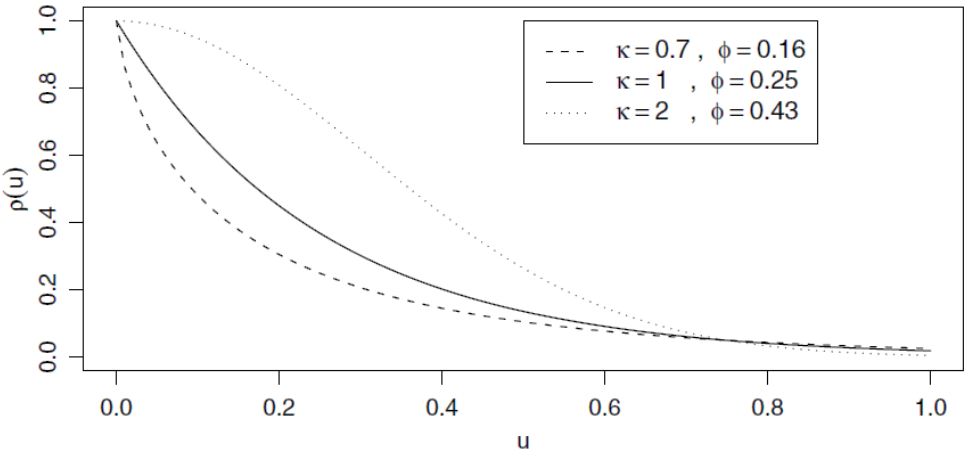


Figure 61: The degree of smoothness in the underlying spatial process associated with different values of the shape parameter, κ, in the Matern family of correlation functions (adapted from Diggle & Ribeiro 2007).

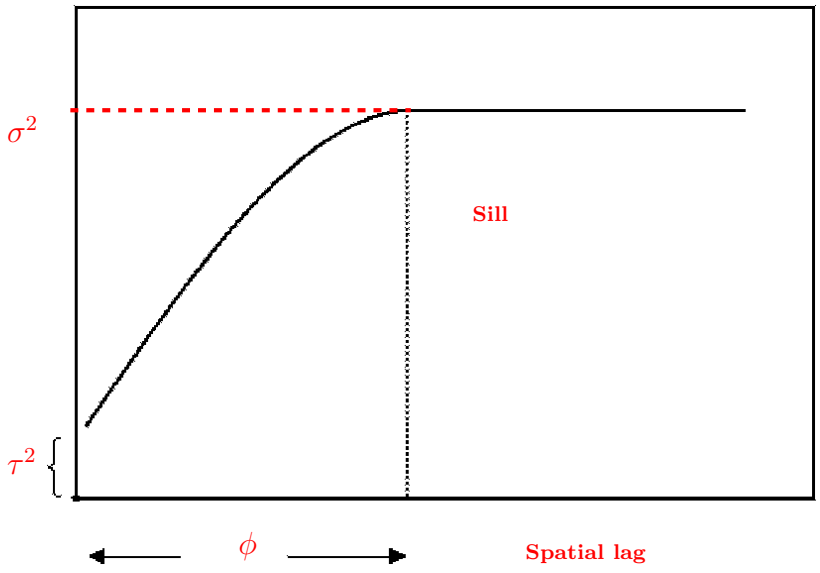


Figure 62: An example of a typical isotropic variogram showing the elements of the covariance structure, ϕ, σ^2 and τ^2. The range, ϕ, quantifies the zone within which attributes are spatially correlated, the variance of the spatial process, Y (s), is the sum of the measurement error variance, τ^2, and signal variance, σ^2. This figure was adapted from Waller & Gotway 2004 with permission of the rights holder, John Wiley and Sons.

2.8.4.2.1 Parameter Re-Estimation and Spatial Prediction

The parameters of our provisional model, together with the estimated parameters of the covariance structure in section 2.8.4.2, were incorporated into a spatial model (equation 14). The regression parameters were re-estimated in the spatial model by an iterative scheme involving the re-estimation of the covariance function for the residuals of each fitted model for 10,000 iterations till the convergence of the correlation parameters. The model, equation 14, was executed in `R` using specific functions as implemented with the `R PrevMap` package by Giorgi & Diggle 2014.

$$\log\left(\frac{p_i}{1-p_i}\right) \quad = \quad d'\beta + S(x) + Z \tag{14}$$

where d' denotes the vector of covariates from the provisional model; β is the vector of regression coefficients; $S(x)$ is a Gaussian process that is described by the covariance parameters; and Z is a mutually independent zero-mean Gaussian variable with variance, τ^2 (Giorgi & Diggle 2014).

The parameter estimation was effected by running a single chain sampler for 10000 iterations. After allowing a suitable burn-in of 2000 iterations, every 8th iteration was stored. The stored values were used in approximating the likelihood integral. To effect the spatial predictions, the prediction locations were first specified by 0.01×0.01 degrees grid. The parameters from binomial spatial model, equation 14, were used in generating the risk maps after summarising the predictive distribution of the spatial grids by their means, standard deviations and probability distribution (Giorgi & Diggle 2014).

3 Results

3.1 Baseline Characteristics

3.1.1 Prevalence of Our Study Outcomes

All the recorded infections in this study occurred in the peri-urban and rural sites (Table 37). Schistosome infections were also observed among both the peri-urban and rural participants. However, *S. mansoni* was the least occurring of the two schistosome infections. Moreover, of the soil-transmitted helminths infections, only hookworm and whipworms were observed among the peri-urban participants. Based on these results, Table 37, we chose to revise our study outcomes from 3 to 2. Therefore, the analysis in the rest of this section would focus on *S. haematobium* mono-infections and the soil-transmitted helminth infections.

Table 37: Recorded infections among participants in the three study sites. The total number of participants in the urban, peri-urban and rural sites were 228, 532 and 579, respectively.

Helminth Infections	Number of Infected Participants			
	Urban (%)	Peri-Urban (%)	Rural (%)	Total
Schistosoma spp				
S. haematobium	0 (0.00)	15 (2.82)	30 (5.18)	45
S. mansoni	0 (0.00)	3 (0.56)	2 (0.35)	5
Soil-Transmitted Helminths (STH)				
Hookworms	0 (0.00)	9 (1.69)	78 (13.47)	87
Roundworms	0 (0.00)	0 (0.00)	52 (8.98)	52
Whipworms	0 (0.00)	2 (0.38)	22 (3.80)	24
Any of the STH infections	0 (0.00)	11 (2.07)	138 (23.83)	149

3.1.2 Socio-Economic Standards of Households

3.1.2.1 Source of Water and Sanitation Facility

Tap water was the major source of water supply for the urban and peri-urban participants while the rural participants relied mainly on surface water sources (Table 38). Water

from tanker trucks was, however, the second most widely used source of water in all the three sites. Even though the vast majority of the urban and peri-urban participants had access to private flush toilets and compound latrines, most participants (77%) in the rural site practised open sanitation. Moreover, there were some few people (4%) in the peri-urban site who also resorted to open sanitation.

3.1.2.2 Type of House, Electricity and Transportation

Compound houses constituted the most common dwellings in all three sites and the main building material for most houses was cement. While most participants came from homes that had a supply of electricity, there were some houses in all the three sites that had no electricity supply (Table 38). Almost all of our rural participants and about a half of all the participants in the urban and peri-urban sites were mainly in the habit of walking to school.

3.1.2.3 Educational Levels of Family Providers

The educational levels of family providers in the urban site were almost equally distributed between basic, secondary, further and higher education. However, basic and higher education were the most common among family providers in the peri-urban site while those in the rural site were mainly with some level of basic education or without any form of formal education (Table 38).

3.1.2.4 Health-Seeking Habits

Adherence to regular treatment routines with over-the-counter anthelmintics was mostly common in the urban and peri-urban homes. Most of our urban (49%) and peri-urban (40%) participants had been treated by their parents in the three months prior to our questionnaire survey. Though there were some level of adherence in the rural site, most parents (44%) were not in the habit of following any treatment routines (Table 38).

Table 38: Indicators of socio-economic standards as well as risk and health-seeking behaviours of our study population versus prevalence of infection.

Indicator Variables	Number of Participants				Prevalence of Infections	
	Urban (%)	Peri-Urban (%)	Rural (%)	Total	*S. haematobium* (%)	STH Infections (%)
Source of Water Supply						
Tap	202 (88.60)	320 (60.15)	3 (0.52)	525	10 (1.90)	9 (1.71)
Tanker	16 (7.02)	138 (25.94)	213 (36.79)	367	13 (3.54)	62 (16.89)
Well	8 (3.51)	71 (13.35)	101 (17.44)	180	10 (5.56)	25 (13.89)
Surface Water	0 (0.00)	1 (0.19)	262 (45.25)	263	12 (4.56)	53 (20.15)
Type of Sanitation Facility						
Private Water Closet	115 (50.44)	222 (41.73)	7 (1.21)	344	3 (0.87)	1 (0.29)
Compound Latrine	91 (39.91)	225 (42.29)	29 (5.01)	345	11 (3.18)	11 (3.19)
Public Latrine	19 (8.33)	59 (11.09)	96 (16.58)	174	5 (2.87)	21 (12.07)
Open Sanitation	0 (0.00)	22 (4.14)	445 (76.86)	467	26 (5.57)	114 (24.41)
Type of House						
Detached	58 (25.44)	174 (32.71)	143 (24.70)	375	12 (3.20)	35 (9.33)
Semi-detached	41 (17.98)	127 (23.87)	86 (14.85)	254	7 (2.76)	26 (10.25)
Compound house	128 (56.14)	229 (43.05)	350 (60.45)	707	26 (3.68)	88 (12.45)
Main Building Material						
Cement	226 (99.12)	491 (92.29)	404 (69.78)	1121	31 (2.77)	112 (9.99)
Mud	2 (0.88)	39 (7.33)	175 (30.22)	216	14 (6.48)	37 (17.13)
Source of Cooking Fuel						
Charcoal	51 (22.37)	162 (30.45)	225 (38.86)	438	19 (4.34)	59 (13.47)
Firewood	0 (0.00)	40 (7.52)	343 (59.24)	383	23 (6.01)	85 (22.19)
Liquified petroleum gas	177 (77.63)	328 (61.65)	9 (1.55)	514	3 (0.58)	5 (0.97)
Electricity Supply in the Home						
No	2 (0.88)	50 (9.40)	144 (24.87)	196	15 (7.65)	39 (19.90)
Yes	226 (99.12)	482 (90.60)	435 (75.13)	1143	30 (2.62)	110 (9.62)
Means of Transportation						
Walk	114 (50.00)	269 (50.56)	577 (99.65)	960	41 (4.27)	147 (15.31)
Bus	51 (22.37)	129 (24.25)	1 (0.17)	181	3 (1.66)	1 (0.55)
Private car	63 (27.63)	132 (24.81)	1 (0.17)	196	1 (0.51)	1 (0.51)
Family Provider's Educational Level						
No formal education	16 (7.02)	39 (7.33)	261 (45.08)	316	23 (7.28)	77 (24.37)
Basic	58 (25.44)	202 (37.97)	270 (46.63)	530	18 (3.40)	61 (11.51)
Secondary	63 (27.63)	99 (18.61)	20 (3.45)	182	3 (1.65)	5 (2.75)
Further	41 (17.98)	80 (15.04)	19 (3.28)	140	1 (0.71)	3 (2.14)
Higher	50 (21.93)	111 (20.86)	5 (0.86)	166	0 (0.00)	2 (1.20)
Period Since Last Anthelminthic Treatment[a]						
< 1 month	21 (9.21)	68 (12.78)	18 (3.11)	107	2 (1.87)	6 (5.61)
1-3 months	112 (49.12)	213 (40.04)	54 (9.33)	379	6 (1.58)	12 (3.17)
3-6 months	56 (24.56)	148 (27.82)	75 (12.95)	279	5 (1.79)	15 (5.38)
> 6 months	31 (13.60)	87 (16.35)	174 (30.05)	292	14 (4.79)	43 (14.73)
Never	3 (1.32)	7 (1.32)	256 (44.21)	266	17 (6.39)	72 (27.07)
Name of Anthelmintic						
Albendazole / Mebendazole	5 (2.19)	72 (13.53)	61 (10.54)	138	5 (0.58)	9 (1.05)
Wormplex	113 (49.56)	253 (47.56)	489 (84.46)	855	36 (26.09)	129 (93.48)
Vermox / Zentel	91 (39.91)	157 (29.51)	6 (1.04)	254	3 (1.18)	4 (1.57)
Other[b]	16 (7.02)	36 (6.77)	14 (2.42)	66	1 (1.52)	4 (6.06)

[a] Refers to home-administered treatment with over-the-counter anthelmintics.

[b] The less commonly used medications such as herbal preparations and milk of magnesia.

3.1.2.5 Types of Anthelmintics

Judging by the types of anthelmintics our respondents used for their regular treatment routines (Table 38), health-seeking habits at the household level were not extended to schistosomiasis treatment. The commonly used anthelmintics were mostly known to be effective against the soil-transmitted helminths but not *Schistosoma spp.* However, some of the respondents also reported the use of herbal medicines whose efficacy levels are either undocumented or unknown.

3.1.3 Prevalence of Infection by Socio-Economic Standards of Households

The soil-transmitted helminth infections were mainly common among participants who came from households that resided in compound houses and whose family providers were either without any formal education or have had some level of basic education. Such households mainly relied on surface water sources and practised open sanitation (Table 38). Moreover, the STH infections tend to be more common among children who came from households that rarely or never observed any treatment routines. In households where treatment regimes were followed, those that administered wormplex tend to have more infected children.

S. haematobium infections, on the other hand, were more common among participants whose households mainly relied on water from wells and practised open sanitation (Table 38). Moreover, these participants mostly resided in compound houses that were built with mud and had no supply of electricity. Infections were also mostly observed among participants who came from households that rarely followed any regular treatment routines (Table 38).

3.1.4 Assessing Non-Linearity in Age

3.1.4.1 Soil-Transmitted Helminth Infections

An initial assessment of the relationship between age and the prevalence of STH infections indicates no infections were observed in children under the age of 6 (Figure 63(a)). To help discern the nature of the trend across the age range of our study population, we

employed the use of non-parametric smoothers by Wood 2006. The smoother for age-STH relationship, Figure 63(b), estimated a significant ($p<0.001$) non-linear effect in age with 8.362 effective degrees of freedom as the optimum amount of smoothing. However, age only explained six percent of the total variation in STH infections (Table 39).

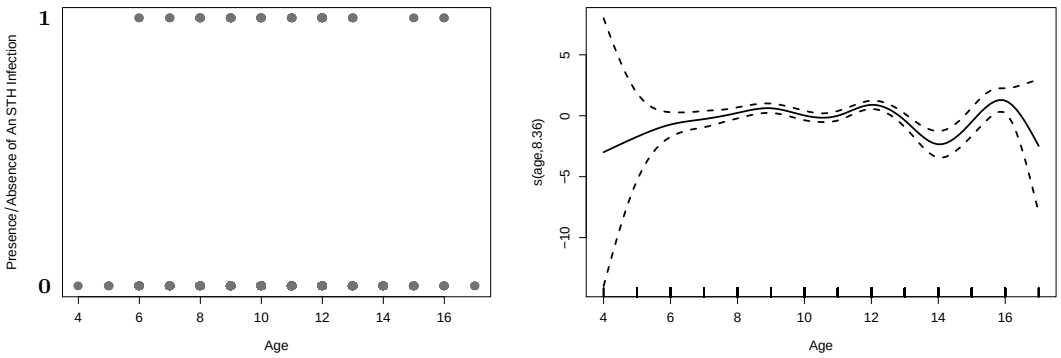


Figure 63: Left: Scatterplot of the prevalence of STH infections in the three sites versus age of the study participants. Right: Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines).

3.1.4.2 *S. haematobium* Mono-Infection

Figure 64(a) indicates that all the observed *S. haematobium* mono-infections occurred in children above 6 years of age. The smoother estimated a non-linear effect with 3.302 effective degrees of freedom (Table 39). However, this effect was non-significant at the 5% level.

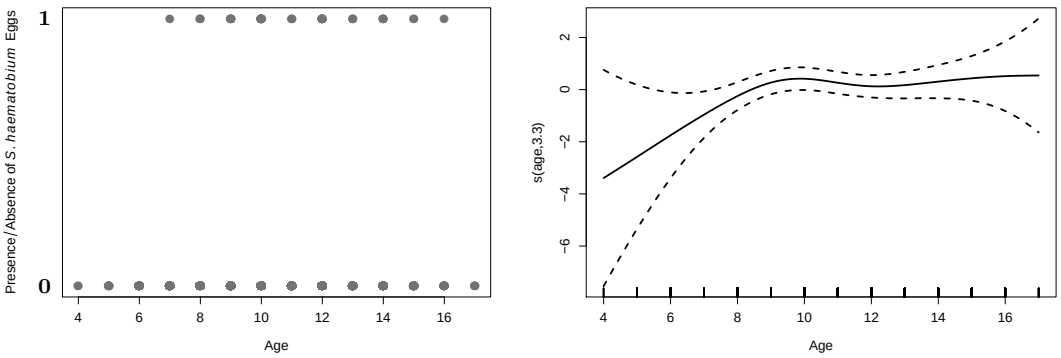


Figure 64: Left: Scatterplot of the prevalence of *S. haematobium* mono-infections in the three sites versus age of the study participants. Right: Estimated non-parametric smoother for the effect of age on the predictor scale (solid line) and the corresponding 95% confidence limits (outer dashed lines).

Table 39: Model summary for the smooth function of age in the generalised additive models (equation 1). By default, the *gam* function in the *mgcv* package used the Unbiased Risk Estimator (UBRE) score in the cross-validation process to determine the optimum amount of smoothing. The amount of smoothing is expressed as effective degrees of freedom (EDF) (Zuur et al. 2011).

	Approximate Significance of Smooth Terms					Fit Details				
Outcome	**Smooth Terms**	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance expl.**	**Scale est.**	**N**
STH	s (Age)	8.362	8.833	39.84	0.000	0.032	-0.281	5.79%	1	1209
S. haematobium	s (Age)	3.302	4.131	5.857	0.223	0.004	-0.701	2.66%	1	1302

3.1.4.2.1 Modelling the Effect of Age

In line with the discussion in section 3.1.2.1.2 in Part I of this thesis, the smoother, $f(\mathrm{X}_i)$; is made up of basic units as given by equation 15. Therefore, when the evidence presented by the estimated non-parametric smoothers and their numerical outputs pointed to non-linearity in the effect of age, the smoothers were regarded as sticks that could be broken down into a given number of segments, p. Hence, by adopting White et al.'s broken stick model approach, the length of each segment of the smoother, j, was defined by equation 16.

$$f(\mathrm{X}_i) = \sum_{j=1}^{p} \beta_j \times b_j(\mathrm{X}_i) \tag{15}$$

where p is the total number of segments that make up $f(\mathrm{X}_i)$ and β_j is the regression coefficient obtained by fitting the j^{th} segment in the model.

$$\mathrm{X_b} = \begin{cases} 0, & \text{if} \quad \mathrm{k_{b-1}} > \mathrm{X}_i \\ \mathrm{X}_i - \mathrm{k_{b-1}}, & \text{if} \quad k_{b-1} \quad \leq \mathrm{X}_i < \mathrm{k}_b \\ \mathrm{k_b}, & \text{if} \quad \mathrm{X}_i > \mathrm{k_b} \end{cases} \tag{16}$$

where $\mathrm{X_b}$ is a given segment of the smoother; $\mathrm{k_{b-1}}$ and $\mathrm{k_b}$ are break-points of the smoother corresponding to points on the X-gradient such that $\mathrm{k_{b-1}} < \mathrm{k_b}$. So, for instance, age is set to 0 when age is less than the first break-point, $\mathrm{k_{b-1}}$.

3.2 Confirmatory Analysis

3.2.1 Modelling the Risk of the Soil-Transmitted Helminths Infections

3.2.1.1 Simple Logistic Regression Analyses

The results of the simple regression analyses for the STH infections are presented in Table 40 below. With the exception of age and type of house; the effect of all the other covariates were highly significant in predicting the risk of the soil-transmitted helminth infections in school-aged children.

3.2.1.2 Multiple Logistic Regression Analysis

The deviance-based test for assessing the significance of the terms in the multiple regression model, Table 41, indicated period since last anthelmintic treatment, age5, electricity supply in the home, educational level of family provider and building material could be excluded from the model without causing any statistically significant increase in the overall deviance.

3.2.1.3 Investigating the Effects of Interaction Terms

Our assessment of interactions between terms in the main effects model yielded one marginally significant effect, fuel:water (Table 42).

3.2.1.4 Provisional Model for the STH Infections

Table 43 presents our provisional model for the STH infections. The inclusion of the fuel:water interaction term in the model caused fitting problems. We, therefore, carried out the rest of the analysis without that term.

Table 40: Simple and multiple regression models for the risk of the soil-transmitted helminth infections.

Model Terms	Simple Regression Model			Deviance Analysis			Multiple Regression Model		
	OR (95% CI)	SE	p-value	D_0 - D_1	df	p-value	OR (95% CI)	SE	p-value
Type of Sanitation Facility									
Private Water Closet	1						1		
Compound Latrine	10.89 (2.10, 1.99 x 10^2)	1.048	0.023	173.22	3	0.000	4.17 (0.70, 80.16)	1.095	0.193
Public Latrine	47.91 (9.87, 8.63 x 10^2)	1.029	0.000				8.70 (1.39, 1.72 x 10^2)	1.122	0.054
Open Sanitation	1.17 x 10^2 (26.01, 2.07 x 10^3)	1.008	0.000				12.09 (1.99, 2.38 x 10^2)	1.114	0.025
Source of Cooking Fuel									
Charcoal	1						1		
Firewood	1.87 (1.29, 2.71)	0.188	0.001	127.19	2	0.000	1.04 (0.68, 1.60)	0.218	0.848
Liquefied petroleum gas	0.06 (0.02, 0.14)	0.471	0.000				0.34 (0.07, 1.24)	0.717	0.134
Family Provider's Educational Level									
No formal education	1						1		
Basic	0.4 (0.27, 0.58)	0.191	0.000	106.2	4	0.000	0.76 (0.50, 1.15)	0.213	0.190
Secondary	0.09 (0.03, 0.20)	0.474	0.000				0.54 (0.17, 1.47)	0.540	0.258
Further	0.06 (0.02, 0.18)	0.599	0.000				0.42 (0.06, 1.61)	0.782	0.271
Higher	0.04 (0.01, 0.12)	0.724	0.000				0.78 (0.11, 3.64)	0.858	0.774
Source of Water Supply									
Tap	1						1		
Tanker	12.64 (0.01, 0.03)	0.365	0.000	105.94	3	0.000	3.29 (1.38, 8.61)	0.463	0.010
Well	9.81 (4.63, 22.70)	0.401	0.000				2.16 (0.83, 6.00)	0.499	0.121
Surface Water	15.27 (7.75, 33.70)	0.371	0.000				2.10 (0.84, 5.76)	0.487	0.127
Period Since Last Anthelmintic Treatment									
< 1 month	1						1		
1-3 months	0.56 (0.21, 1.65)	0.513	0.261	101.52	4	0.000	0.89 (0.30, 2.86)	0.564	0.840
3-6 months	1.02 (0.40, 2.94)	0.498	0.966				1.01 (0.36, 3.14)	0.543	0.980
> 6 months	3.24 (1.43, 8.72)	0.453	0.010				1.50 (0.59, 4.37)	0.505	0.423
Never	6.3 (2.85, 16.72)	0.444	0.000				1.68 (0.67, 4.86)	0.499	0.299
Means of Transportation									
Walk	1						1		
Bus	0.03 (0.00, 0.15)	1.007	0.001	86.26	2	0.000	0.00 (0.00, 0.00)	489.320	0.976
Private car	0.03 (0.00, 0.12)	1.006	0.000				0.37 (0.02, 2.19)	1.092	0.364
Age (years)									
age1	4.50 x 10^5 (0, NA)	408.430	0.975	2.38	1	0.123	-		
age2	1.30 (1.03, 1.69)	0.124	0.034	5.15	1	0.023	1.23 (0.93, 1.66)	0.146	0.156
age3	1.03 (0.86, 1.25)	0.097	0.729	0.12	1	0.729	-		
age4	1.29 (0.91, 1.83)	0.178	0.146	2.09	1	0.148	-		
age5	0.60 (0.41, 0.83)	0.177	0.004	23.63	1	0.001	1.20 (0.76, 1.84)	0.224	0.416
age6	1.26 (0.74, 1.98)	0.247	0.344	0.83	1	0.362	-		
age7	0.00 (NA, x 10^{16})	441.372	0.977	1.06	1	0.303	-		
Electricity Supply in the Home									
No	1						1		
Yes	0.43 (0.29, 0.65)	0.207	0.000	15.15	1	0.000	0.88 (0.56, 1.39)	0.231	0.571
Main Building Material									
Cement	1						1		
Mud	1.9 (1.25, 2.83)	0.208	0.002	8.72	1	0.003	0.64 (0.40, 1.01)	0.233	0.058
Sex									
Female	1						1		
Male	1.47 (1.04, 2.08)	0.176	0.028	4.88	1	0.027	1.35 (0.92, 1.99)	0.198	0.131
Type of House									
Detached	1								
Semi-detached	1.08 (0.62, 1.84)	0.274	0.786	3.42	2	0.181	-		
Compound house	1.42 (0.95, 2.18)	0.212	0.095						

- : Indicates the corresponding term did not qualify for inclusion in the multiple regression model.

Table 41: Deviance analysis for terms in the multiple regression model for the soil-transmitted helminth infections. The p-values were obtained by the sequential omission of the model terms, starting with the least significant.

	Deviance Analysis				
Model Terms	*df*	D_0 - D_1	**Residual Deviance**	**Residual** *df*	**p-value**
Type of Sanitation Facility					
Private Water Closet					
Compound Latrine	3	167.21	711.40	1173	0.000
Public Latrine					
Open Sanitation					
Source of Cooking Fuel					
Charcoal					
Firewood	2	14.48	696.93	1171	0.001
Liquified petroleum gas					
Family Provider's Educational Level					
No formal education					
Basic	4	7.99	688.94	1167	0.092
Secondary					
Further					
Higher					
Source of Water Supply					
Tap					
Tanker	3	13.47	675.47	1164	0.004
Well					
Surface Water					
Period Since Last Anthelminthic Treatment					
$<$ 1 month					
1-3 months	4	4.99	670.47	1160	0.288
3-6 months					
$>$ 6 months					
Never					
Means of Transportation					
Walk					
Bus	2	9.03	661.44	1158	0.011
Private car					
Age (years)					
age5	1	0.78	660.67	1157	0.378
Electricity Supply in the Home					
No					
Yes	1	0.05	660.61	1156	0.820
Main Building Material					
Cement					
Mud	1	3.60	657.02	1155	0.058
Age (years)					
age2	1	2.54	654.48	1154	0.111
Sex					
Female					
Male	1	2.29	652.19	1153	0.131

Table 42: Deviance analysis for interaction terms in the main effects model for the STH infections. D_0 - D_1 represents the differences between the null and residual deviances.

	Deviance Analysis				
Interaction Terms	*df*	D_0 - D_1	**Residual Deviance**	**Residual *df***	**p-value**
Type of Sanitation Facility					
Sanitation : Source of Fuel	6	9.890	669.65	1178	0.129
Sanitation: Water	9	8.585	670.95	1175	0.476
Sanitation: Transport	5	4.801	674.74	1179	0.441
Source of Cooking Fuel					
Source of Fuel : Water	5	10.90	668.64	1179	0.053
Source of Fuel : Transport	4	1.159	678.38	1180	0.885
Source of Water in the Home					
Source of Water : Transport	5	2.550	676.99	1179	0.769

Table 43: Provisional model for the soil-transmitted helminth infections.

	Provisional Model		
Model Terms	**OR (95% CI)**	**SE**	**p-value**
Type of Sanitation Facility			
Private Water Closet	1		
Compound Latrine	5.32 (0.94, 1.0 x10^2)	1.079	0.121
Public Latrine	11.68 (2.02, 2.24 x 10^2)	1.092	0.024
Open Sanitation	17.91 (3.20, 3.40 x10^2)	1.082	0.008
Source of Cooking Fuel			
Charcoal	1		
Firewood	1.12 (0.74, 1.68)	0.208	0.600
Liquified petroleum gas	0.36 (0.10, 1.08)	0.599	0.092
Source of Water Supply			
Tap	1		
Tanker	3.96 (1.77, 9.79)	0.431	0.001
Well	2.44 (1.00, 6.44)	0.471	0.059
Surface Water	2.32 (1.00, 5.94)	0.450	0.062
Means of Transportation			
Walk	1		
Bus	0.00 (0.00, 0.00)	486.783	0.975
Private car	0.30 (0.02, 1.69)	1.079	0.261

3.2.1.5 Goodness-Of-Fit Check

The value of the dispersion statistic, 0.57, fell without our acceptable cut-range as discussed in section 2.8.2.1. Based on this result, we concluded the model was not over-dispersed.

3.2.2 Model Validation

3.2.2.1 Assessing the Homoscedasticity Assumption

The spread of the residuals across the range of the fitted values seems reasonably constant. This, therefore, ruled out any obvious violation of the homoscedasticity assumption (Figure 65). Hence, the provisional model was deemed acceptable.

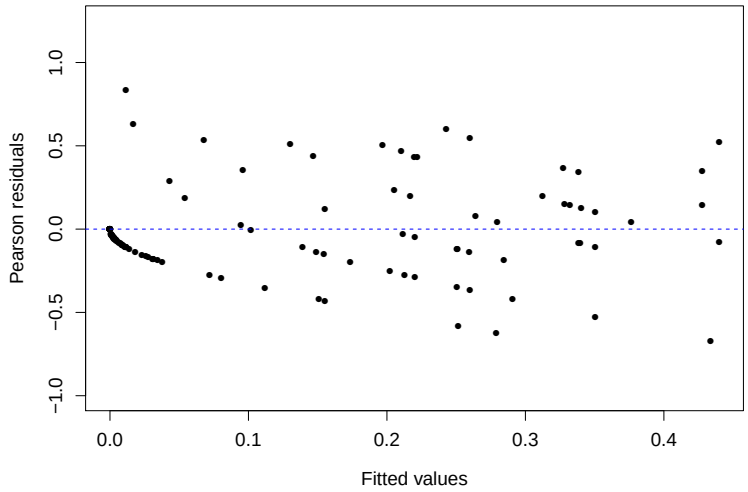


Figure 65: Grouped Pearson residuals versus fitted values of the provisional model for the soil-transmitted helminths (STH) infections. The spread of residuals across the range of the fitted values suggests homoscedasticity is a reasonable assumption.

3.2.2.2 Assessing the Assumption of Independence

Figure 66 mainly shows a segregation of negative and positive residuals in the urban and peri-urban sites. A comparison with Figure 67 suggests our provisional model mainly over-estimated the risk of infection in the urban and peri-urban sites. The rural sites, however, did not seem to show any clear segregation of risk as judged by the point map, Figure 66. In the next section, we employ the use of the variogram as a more objective tool for quantifying the spatial dependence in the residuals.

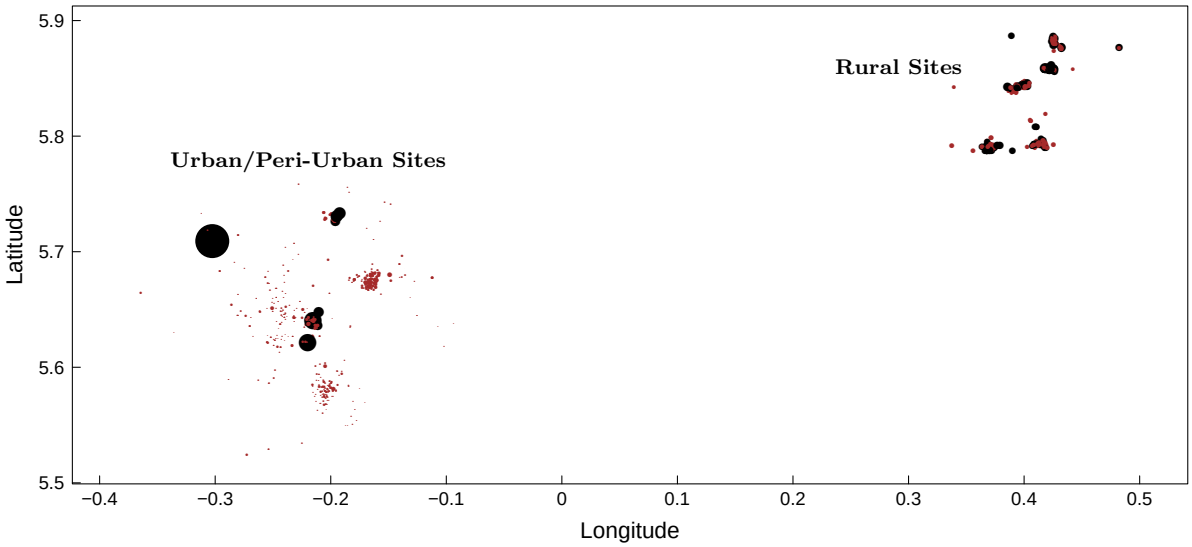


Figure 66: Spatially-referenced residuals of the provisional model for STH infections in the three sites (urban/peri-urban and rural). The radii of the circles are proportional to the absolute values of the Pearson residuals. The red and black circles represent the negative and positive residuals, respectively.

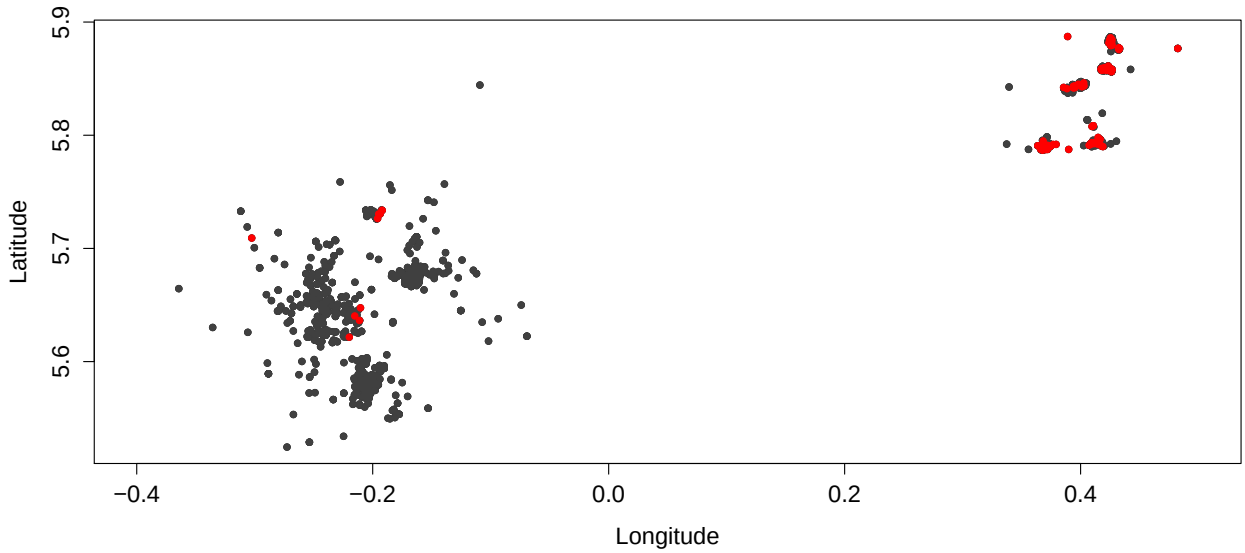


Figure 67: Distribution of STH infections in the three sites. Each point corresponds to a sampled location i.e location of the household of the sampled child. The red and black points represent infected and non-infected cases, respectively.

3.2.2.2.1 Structural Analysis

The omni-directional empirical variogram suggests an increasing spatial trend with lag distance (Figure 68 (A)). However, this variogram is based on the isotropy assumption where the spatial correlation is thought to be identical in all directions within the study region (Waller & Gotway 2004). Since isotropy may not be a reasonable assumption for the outcome under consideration, we conducted an assessment for directional dependence. But first, we examine some possible sources of violation of the isotropy assumption in the context of this study.

3.2.2.2.2 Violation of the Isotropy Assumption

The basis for modelling both the STH and schistosome infections on the same set of covariates is that the two diseases tend to be co-endemic; thereby suggesting that both diseases may be favoured by similar conditions (Gazzinelli et al. 2012, Lustigman et al. 2012). Moreover, chemotherapy interventions for these infections are often implemented concurrently at the same geographical unit and among the same host population (Brooker et al. 2009)

However, while schistosome infections may be unevenly distributed within any given endemic setting, thereby suggesting a possible violation of the isotropy assumption (Brooker 2007), the STH infections are not generally known to exhibit any such patterns in space. However, the fact that the transmission of both diseases are favoured by similar factors implies that they could share similar patterns in their distribution.

3.2.2.2.3 Assessing Directional Dependence

Directional empirical variograms were computed for four azimuths in our study region. The horizontal bands for the 45 and 0 degrees angles in Figure 68 (B), suggest the absence of any significant spatial dependency in those directions. For the 90 and 145 degrees angles, however, the directional variograms seemed to have different ranges, which could signify geometric anisotropy (Waller & Gotway 2004).

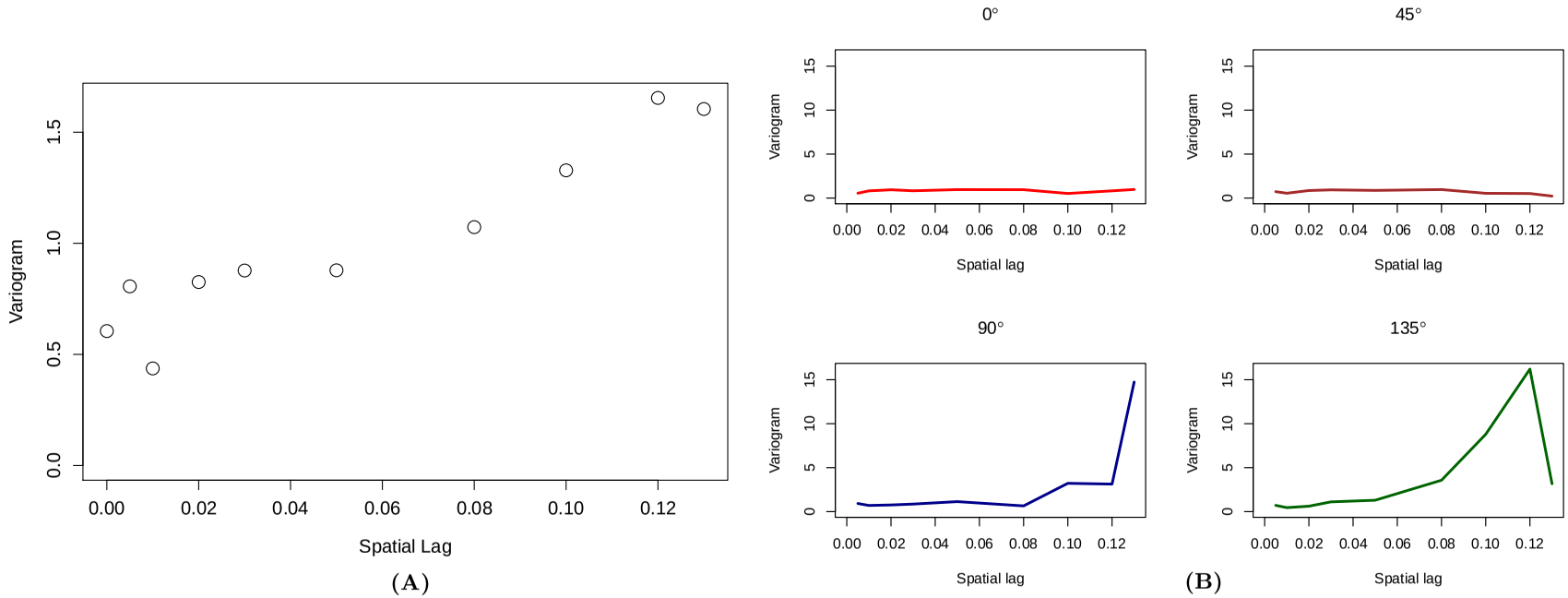


Figure 68: (A) Omni-directional empirical variogram for the Pearson residuals of the provisional model for the STH infections. (B) Directional empirical variograms for lags in the 0°, 45°, 90° and 135° spatial directions, with a tolerance angle of ± 22.5°. The slopes of the directional variograms suggest isotropy may not be a reasonable assumption in this case.

The angle of maximum range and the range ratio, Figure 69, were estimated as 30.175 degrees and 1.20; respectively using specific functions in the `R intamap` package (Pebesma et al. 2010). Therefore, the geographical coordinates were transformed to correct for the anisotropy using the estimated anisotropy parameters (Ribeiro Jr & Diggle 2001). The isotropic variogram, that was computed with the anisotropy-corrected coordinates, was statistically significant, p=0.038, for spatial autocorrelation. Hence, the rest our assessments will be based on this isotropic variogram.

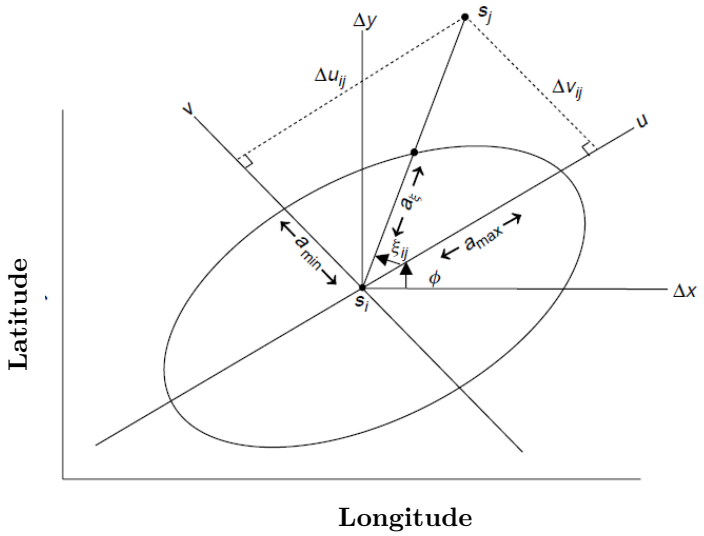


Figure 69: A schematic representation of geometric anisotropy. a_{max} denotes the range of maximum continuity in direction φ; and $a_{min}$is the range in the direction of minimum continuity perpendicular to the angle of maximum range (adopted from Waller & Gotway 2004 with permission of the rights holder, John Wiley and Sons.)

3.2.2.2.4 Estimating the Covariance Parameters

Figure 70 shows the fitted Maťern covariance function, equation 13, that was used in modelling the spatial correlation structure. The parameters of the covariance structure, given by the range of spatial correlation, the signal variance and the measurement error variance, were estimated as $\phi = 1.78$, $\sigma^2 = 4415.24$ and $\tau^2 = 0.55$; respectively by weighted least squares.

The shape of the fitted parametric model is, however, consistent with an unbounded variogram. This may occur as result of increasing sources of variation with lag distance and cannot be modelled with a finite sill and range (Thangarajan 2007, Webster & Oliver

2001). Therefore, to adjust for the high variance in Figure 70, the range and sill of the covariance function were set at 0.08° and 2.0, respectively.

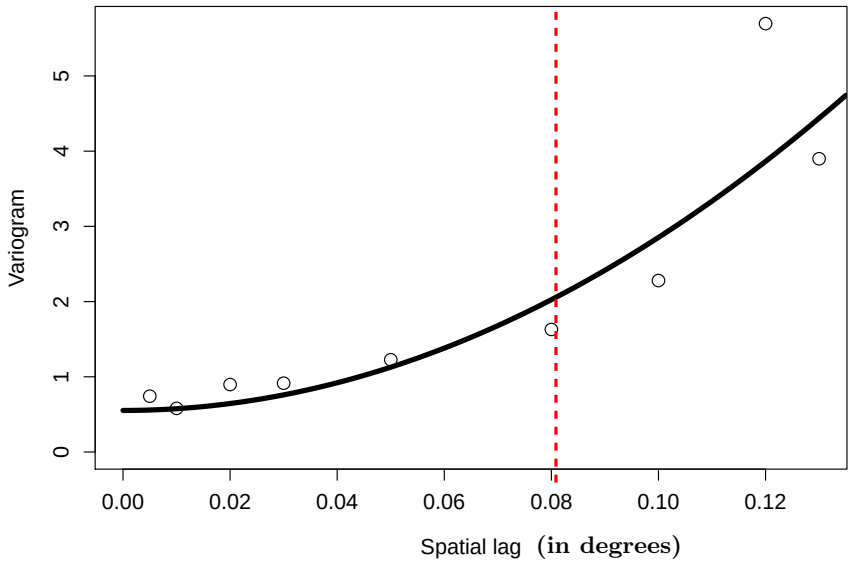


Figure 70: The isotropic variogram for the STH infections with the fitted covariance function of the order $\kappa = 2.5$. The covariance parameters were estimated as: $\phi = 1.78$, $\tau^2 = 0.55$, $\sigma^2 = 4415.24$. To adjust for the high variance, ϕ was set at 0.08°; as indicated by the vertical dashed-line.

3.2.2.2.5 Geostatistical Model

Our final model is presented in Table 44. A comparison of our final and provisional models for STH infection indicates that the provisional model (Table 43), which did not take the spatial correlation into account, over-estimated the covariate effects due to lack of independent information on the autocorrelated data.

3.2.2.2.6 Predicted Prevalence of the STH Infections

The predicted prevalence of STH infections, based on the covariates in our final model (Table 44), mostly shows a low prevalence of STH infections across the region (Figure 71). The standard errors for these prediction were highest for those areas with the slightly higher predicted prevalence and vice versa (Figure 72).

Table 44: Regression parameters and standard errors of the final and provisional models for STH infections.

	Geostatistical Model			Provisional Model		
Parameters	**Posterior Mean**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Private Water Closet	1					
Compound Latrine	1.589	1.088	0.144	1.6714	1.079	0.121
Public Latrine	2.069	1.120	0.065	2.4576	1.092	0.024
Open Sanitation	2.447	1.112	0.028	2.8852	1.082	0.008
Charcoal	1					
Firewood	0.097	0.222	0.663	0.109	0.208	0.600
Liquefied petroleum gas	-1.228	0.692	0.076	-1.011	0.599	0.092
Tap	1					
Tanker	0.605	0.484	0.212	1.375	0.431	0.001
Well	0.314	0.525	0.550	0.890	0.471	0.059
Surface Water	0.259	0.498	0.602	0.840	0.450	0.062
Walk	1					
Bus	-15.134	816.809	0.985	-15.185	486.783	0.975
Private car	-0.906	1.088	0.405	-1.211	1.079	0.261
Covariance Parameters						
$\log(\sigma^2)$	0.022	1.041				
$\log(\phi)$	-1.934	0.463				
$\log(\tau^2)$	-0.771	2.079				

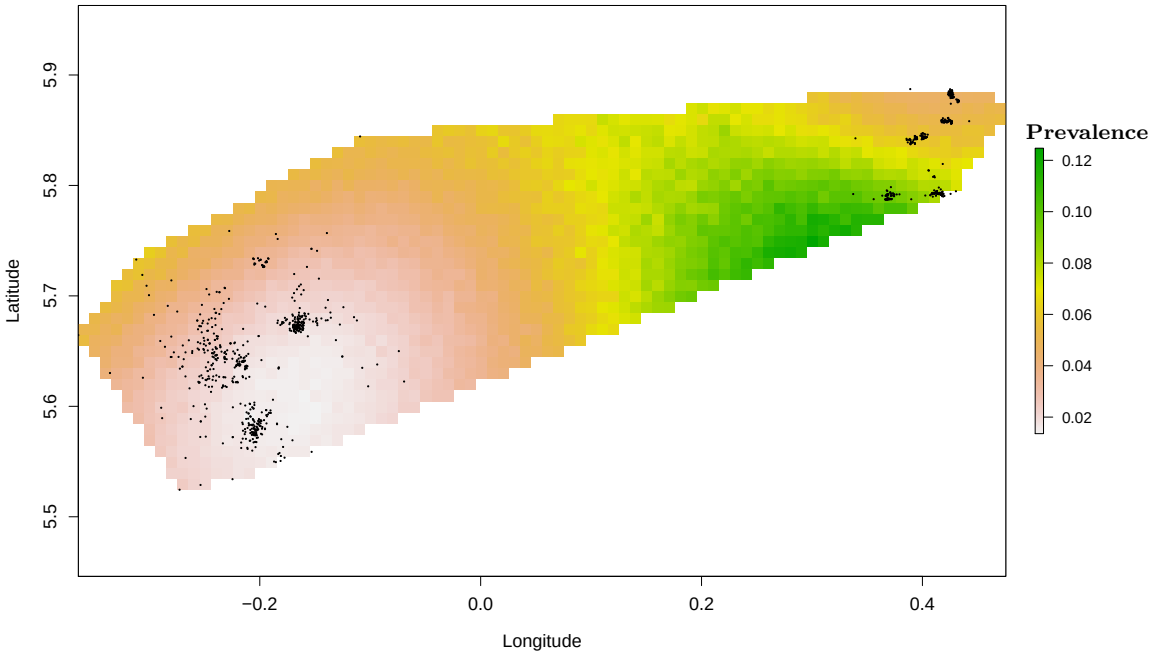


Figure 71: Predicted prevalence of Soil-Transmitted Helminth infections based on the covariates in the geostatistical model (Table 44). The solid black dots represent the locations of the households of the children who supplied data.

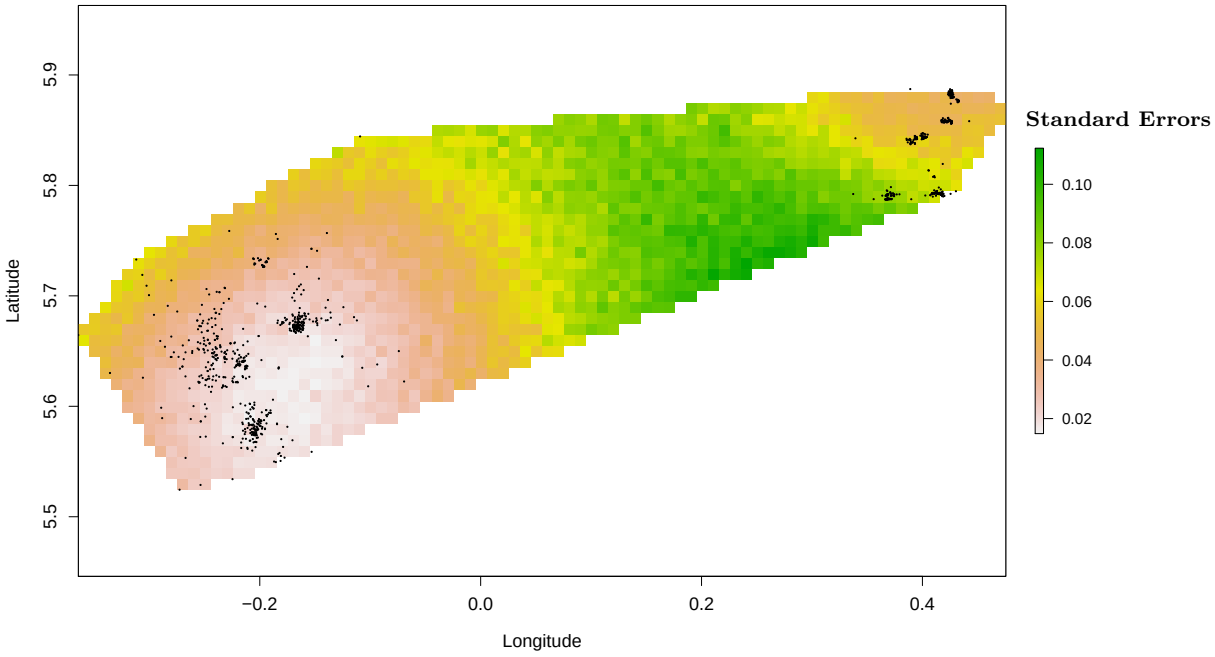


Figure 72: Standard errors associated with the predicted prevalence of Soil-Transmitted Helminth infections in Figure 71

3.2.3 Modelling the Risk of *Schistosoma haematobium* Mono-Infections

3.2.3.1 Mixed Effects Model

Based on the results of the generalised likelihood ratio test in section 2.8.1.6, the risk of *S. haematobium* mono-infections was initially investigated with a random intercept model that had household as the grouping factor (Table 45).

Table 45: Mixed effects models for *S. haematobium* mono-infections.

Model Terms	Fixed Effects			Random Effects		
	Parameter Estimate	SE	p-value	Variance	No. Observations	No. Groups
Family Provider's Educational Level						
No formal education	1					
Basic	-0.485	0.973	0.619	19.581	1298	1075
Secondary	-1.336	2.176	0.539			
Further	-2.054	3.539	0.562			
Higher	-17.049	8017.835	0.998			
Source of Cooking Fuel						
Charcoal	1					
Firewood	-0.266	1.395	0.849	32.105	1298	1075
Liquified petroleum gas	-2.502	3.194	0.434			
Means of Transportation						
Walk	1					
Bus	-0.988	2.171	0.649	20.002	1300	1078
Private car	-1.977	3.536	0.576			
Period Since Last Anthelminthic Treatment						
$<$ 1 month	1					
1-3 months	-1.055	3.607	0.770	38.334	1286	1068
3-6 months	-0.760	3.491	0.828			
$>$ 6 months	0.186	3.302	0.955			
Never	0.823	3.293	0.803			
Type of Sanitation Facility						
Private Water Closet	1					
Compound Latrine	0.343	2.861	0.905	38.241	1293	1070
Public Latrine	0.455	3.089	0.883			
Open Sanitation	1.219	2.658	0.646			
Source of Water Supply						
Tap	1					
Tanker	-0.113	1.766	0.949	37.951	1298	1076
Well	0.580	1.699	0.733			
Surface Water	0.370	1.901	0.846			
Age (years)						
age1	0.487	0.798	0.542	37.857	1302	1078
age2	0.248	0.351	0.480	37.768	1302	1078
age3	0.016	0.088	0.857	25.354	1302	1078
Main Building Material						
Cement	1					
Mud	0.472	1.125	0.675	19.832	1300	1077
Electricity Supply in the Home						
No	1					
Yes	-2.362	1.226	0.054	30.862	1302	1078
Type of House						
Detached	1					
Semi-detached	0.029	1.347	0.983	20.293	1299	1077
Compound house	0.011	1.124	0.992			
Sex						
Female	1					
Male	-0.601	0.912	0.510	37.812	1302	1078

The varying number of observations and group numbers reflect the constant exposure patterns within households.

The results of these analyses, however, suggested that none of the measured covariates were significantly associated with the risk of *S. haematobium* infections (Table 45). However, the standard errors corresponding to the parameter estimates were also high, thereby making the results difficult to interpret (Table 45). In the following sections, we outline the steps that were taken to investigate the adequacy the mixed effects model.

3.2.3.1.1 Random Intercept Model

Building on our earlier concepts in section 2.8.1.5, the random intercept model is given by equation 17 where the model terms have the same meanings as before. The very limited group-level information, as judged by the number of observations versus the group numbers in Table 45, could however render the effect of the random variation negligible. If that were indeed the case, then the random effects model would be expected to break done into a classical fixed effects model. We, therefore, proceeded to investigate if a fixed effects model would have provided an adequate fit as follows. Two fixed effects models were fitted, one to the entire dataset and the other to a subset of the data containing unique observations per household in this study.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) \quad = \quad \beta\mathbf{X}_{ij} \quad + \quad U_i \tag{17}$$

3.2.3.1.2 Rationale for Fitting the Two Fixed Effects Models

If the effect of the group-level homogeneity was not negligible, then the single-level fixed effects model that was fitted to the entire dataset should produce biased parameter estimates and standard errors. However, if the effect of the group-level homogeneity was indeed negligible, then the two models would ideally be expected to have similar parameter estimates and standard errors. The results of these analyses are presented below.

Table 46: Estimated regression parameters and standard errors obtained by fitting a fixed effects logistic regression model to the data.

Model Terms	Simple Regression Model			Deviance Analysis			Multiple Regression Model			Deviance Analysis			
	OR (95 % CI)	SE	p-value	D_0 - D_1	df	p-value	OR (95 % CI)	SE	p-value	D_0 - D_1	Residual Deviance	Residual df	p-value
Family Provider's Educational Level													
No formal education	1						1			27.444	354.78	1263	0.000
Basic	0.46 (0.24, 0.86)	0.323	0.015	29.42	4	0.000	0.60 (0.30, 1.21)	0.357	0.157				
Secondary	0.21 (0.05, 0.61)	0.621	0.012				0.43 (0.09, 1.52)	0.706	0.227				
Further	0.09 (0.01, 0.44)	1.027	0.020				0.19 (0.01, 1.10)	1.083	0.129				
Higher	0 (0.00, 8.62 x 10^8)	839.747	0.984				0 (0.00, 1.14x10^9)	813.942	0.985				
Source of Cooking Fuel													
Charcoal	1						1			7.74	347.04	1261	0.021
Firewood	1.42 (0.76, 2.67)	0.319	0.273	27.31	2	0.000	1.08 (0.52, 2.28)	0.373	0.831				
Liquified petroleum gas	0.13 (0.03, 0.37)	0.625	0.001				0.22 (0.05, 0.82)	0.711	0.035				
Type of Sanitation Facility													
Private Water Closet	1						1			2.20	344.84	1258	0.532
Compound Latrine	3.76 (1.16, 16.72)	0.656	0.044	15.86	3	0.001	1.13 (0.30 , 5.53)	0.718	0.861				
Public Latrine	3.51 (0.85, 17.28)	0.737	0.088				0.48 (0.09, 2.90)	0.857	0.392				
Open Sanitation	6.92 (2.41, 29.20)	0.614	0.002				0.60 (0.13, 3.39)	0.808	0.520				
Period Since Last Anthelminthic Treatment													
< 1 month	1						1						
1-3 months	0.85 (0.19, 5.86)	0.824	0.844	15.80	4	0.003	1.89 (0.40, 13.67)	0.856	0.456	3.29	341.55	1254	0.510
3-6 months	0.99 (0.21, 7.00)	0.845	0.993				1.59 (0.32, 11.64)	0.869	0.593				
> 6 months	2.72 (0.74, 17.49)	0.765	0.191				2.80 (0.69, 19.20)	0.811	0.204				
Never	3.67 (1.03, 23.40)	0.757	0.086				3.11 (0.74, 21.80)	0.825	0.169				
Means of Transportation													
Walk	1						1						
Bus	0.37 (0.09, 1.02)	0.604	0.096	12.32	2	0.002	0.92 (0.20, 3.13)	0.680	0.900	0.41	341.14	1252	0.815
Private car	0.11 (0.01, 0.52)	1.015	0.031				0.52 (0.03, 3.17)	1.101	0.556				
Electricity Supply in the Home													
No	1						1						
Yes	0.32 (0.17, 0.62)	0.327	0.000	10.6	1	0.001	0.64 (0.32, 1.32)	0.359	0.214	1.61	339.53	1251	0.204
Age (years)													
age1	2.20 (1.24, 5.29)	0.357	0.027	8.60	1	0.003	1.97 (1.07, 4.87)	0.374	0.069	4.90	334.63	1250	0.03
age2	1.10 (0.88, 1.40)	0.118	0.400	0.72	1	0.396	-			-			
age3	1.01 (0.96, 1.06)	0.026	0.605	0.26	1	0.610	-			-			
Source of Water Supply													
Tap	1						1						
Tanker	1.92 (0.84, 4.55)	0.426	0.125	8.02	3	0.046	0.83 (0.29, 2.37)	0.533	0.729	0.89	333.74	1247	0.829
Well	3.2 (1.29, 7.94)	0.456	0.011				0.97 (0.31, 2.94)	0.572	0.952				
Surface Water	2.53 (1.08, 6.08)	0.435	0.033				0.64 (0.20, 2.05)	0.594	0.446				
Main Building Material													
Cement	1												
Mud	2.49 (1.26, 4.67)	0.331	0.006	6.69	1	0.010	1.16 (0.54, 2.40)	0.377	0.690	0.16	333.59	1246	0.692
Type of House													
Detached	1												
Semi-detached	0.86 (0.32, 2.18)	0.483	0.762	0.61	2	0.737	-			-			
Compound house	1.18 (0.60, 2.45)	0.355	0.643										
Sex													
Female	1												
Male	1.07 (0.59, 1.94)	0.304	0.829	0.05	1	0.823	-			-			

Table 47: Estimated regression parameters and standard errors obtained by fitting a fixed effects logistic regression model to a subset of the data containing unique observations per household.

Model Terms	Simple Regression Model			Deviance Analysis			Multiple Regression Model			Deviance Analysis			
	OR (95 % CI)	SE	p-value	D_0 - D_1	df	p-value	OR (95 % CI)	SE	p-value	D_0 - D_1	Residual Deviance	Residual df	p-value
Family Provider's Educational Level													
No formal education	1						1						
Basic	0.54 (0.26, 1.08)	0.358	0.081	22.38	4	0.000	0.72 (0.33, 1.57)	0.394	0.413	20.63	285.93	1040	0.000
Secondary	0.18 (0.03, 0.63)	0.755	0.022				0.35 (0.05, 1.55)	0.842	0.213				
Further	0.12 (0.01, 0.61)	1.036	0.043				0.24 (0.01, 1.47)	1.105	0.200				
Higher	0 (0.00, 1.11x10e12)	918.777	0.985				0 (0.00, 4.0x10^5)	892.442	0.986				
Source of Cooking Fuel													
Charcoal	1						1						
Firewood	1.27 (0.63, 2.57)	0.358	0.511	19.68	2	0.000	0.85 (0.38, 1.94)	0.414	0.704	5.22	280.71	1038	0.073
Liquified petroleum gas	0.15 (0.03, 0.44)	0.633	0.002				0.29 (0.06, 1.15)	0.748	0.095				
Means of Transportation													
Walk	1						1						
Bus	0.14 (0.01, 0.66)	1.018	0.054	13.14	2	0.001	0.28 (0.01, 1.52)	1.066	0.230	1.67	279.04	1036	0.434
Private car	0.14 (0.01, 0.64)	1.018	0.051				0.46 (0.02, 2.92)	1.118	0.485				
Period Since Last Anthelminthic Treatment													
< 1 month	1						1						
1-3 months	0.87 (0.20, 6.03)	0.825	0.869	12.53	4	0.014	1.82 (0.38, 1.33)	0.866	0.488	2.31	272.29	1031	0.679
3-6 months	0.61 (0.10, 4.66)	0.922	0.588				0.97 (0.15, 7.86)	0.955	0.975				
> 6 months	2.06 (0.53, 13.58)	0.785	0.357				1.91 (0.44, 1.35)	0.836	0.439				
Never	3.16 (0.86, 20.41)	0.767	0.133				2.61 (0.58, 1.90)	0.853	0.261				
Type of Sanitation Facility													
Private Water Closet	1												
Compound Latrine	2.57 (0.74, 11.84)	0.682	0.166	10.71	3	0.013	1						
Public Latrine	3.29 (0.79, 16.20)	0.738	0.107				0.75 (0.18, 3.84)	0.752	0.707	1.38	270.92	1028	0.711
Open Sanitation	5.33 (1.81, 22.81)	0.625	0.007				0.50 (0.09, 3.08)	0.872	0.428				
							0.47 (0.10, 2.82)	0.833	0.371				
Source of Water Supply													
Tap	1						1						
Tanker	1.96 (0.76, 5.18)	0.481	0.163	7.76	3	0.051	0.90 (0.28, 2.92)	0.597	0.862	1.42	269.50	1025	0.701
Well	3.77 (1.41, 10.23)	0.496	0.008				1.14 (0.33, 3.93)	0.629	0.832				
Surface Water	2.43 (0.92, 6.56)	0.493	0.072				0.63 (0.17, 2.36)	0.670	0.487				
Age (years)													
age1	2.03 (1.15, 4.82)	0.351	0.044	6.71	1	0.010	1.84 (1.01, 4.49)	0.368	0.097	4.43	274.60	1035	0.035
age2	1.06 (0.82, 1.38)	0.130	0.636	0.22	1	0.639	-				-		
age3	1.03 (0.97, 1.08)	0.028	0.322	0.92	1	0.337							
Main Building Material													
Cement	1						1						
Mud	2.69 (1.28, 5.38)	0.364	0.007	6.53	1	0.011	1.26 (0.54, 2.81)	0.416	0.577	0.47	269.03	1024	0.494
Electricity Supply in the Home													
No	1						1						
Yes	0.36 (0.18, 0.78)	0.373	0.006	6.45	1	0.011	0.66 (0.30, 1.56)	0.418	0.327	0.92	268.11	1023	0.336
Type of House													
Detached	1												
Semi-detached	0.54 (0.15, 1.62)	0.591	0.302	1.49	2	0.475	-				-		
Compound house	1.00 (0.48, 2.18)	0.379	0.999										
Sex													
Female	1												
Male	1.05 (0.54, 2.05)	0.340	0.881	0.02	1	0.888	-						

Table 48: Deviance analysis for interaction terms in the two fixed effects models for *S. haematobium* mono-infection.

	Deviance Analysis				
Interaction Terms	***df***	**D_0 - D_1**	**Residual Deviance**	**Residual *df***	**p-value**
Fixed effects model fitted to the entire data					
Educational level : Source of Fuel	8	3.33	343.81	1279	0.912
Educational level : age1	4	1.02	346.12	1283	0.907
Fuel: age1	2	0.66	346.47	1285	0.717
Fixed effects model fitted to the subset of unique observations per household					
Educational level : Source of Fuel	8	2.73	279.87	1054	0.950
Educational level : age1	4	0.74	281.86	1058	0.946
Fuel: age1	2	0.86	281.74	1060	0.652

Table 49: Estimated regression parameters and standard errors in the provisional models obtained by fitting fixed effects logistic regression models to the entire dataset; and to a subset of the data containing unique observations per household

	Entire Dataset			Subset of Unique Observations		
Model Terms	**OR (95 % CI)**	**SE**	**p-value**	**OR (95 % CI)**	**SE**	**p-value**
Family Provider's Educational Level						
No formal education	1			1		
Basic	0.62 (0.28, 1.07)	0.336	0.080	0.63 (0.30, 1.32)	0.372	0.219
Secondary	0.45 (0.09, 1.33)	0.664	0.180	0.32 (0.05, 1.27)	0.800	0.152
Further	0.20 (0.01, 1.00)	1.059	0.116	0.23 (0.01, 1.29)	1.076	0.172
Higher	0.00 (0.00, 5.56 x 10^8)	805.209	0.984	0.00 (0.00, 6.06 x 10^{11})	886.911	0.986
Source of Cooking Fuel						
Charcoal	1			1		
Firewood	1.00 (0.53, 1.97)	0.335	0.971	0.90 (0.43, 1.89)	0.376	0.773
Liquified petroleum gas	0.24 (0.05, 0.72)	0.647	0.024	0.28 (0.06, 0.88)	0.657	0.050
Age						
age1	1.09 (1.12, 5.12)	0.374	0.052	1.87 (1.03, 4.55)	0.365	0.086

3.2.3.1.3 Goodness-of-Fit Check

Both models were under-dispersed with a dispersion statistic of 0.27. Table 50 presents the models estimates from the re-fitted quasi-binomial models. Judging by these results, the two fixed effects models were not materially different from each other.

Table 50: Quasi-binomial model fitted to account for the under-dispersion in the two provisional models in Table 49.

	Entire Dataset			Subset of Unique Observations		
Model Terms	**OR (95 % CI)**	**SE**	**p-value**	**OR (95 % CI)**	**SE**	**p-value**
Family Provider's Educational Level						
No formal education	1			1		
Basic	0.56 (0.31, 1.00)	0.301	0.052	0.63 (0.33, 1.22)	0.334	0.171
Secondary	0.41 (0.11, 1.19)	0.596	0.135	0.32 (0.06, 1.11)	0.719	0.111
Further	0.19 (0.02, 0.87)	0.950	0.080	0.23 (0.02, 1.11)	0.966	0.129
Higher	0.00 (0, 1.34 x10^{7})	722.821	0.983	0.00 (0, 4.37 x10^{8})	796.827	0.984
Source of Cooking Fuel						
Charcoal	1			1		
Firewood	1.01 (0.56, 1.84)	0.301	0.967	0.90 (0.46, 1.75)	0.337	0.748
Liquified petroleum gas	0.23 (0.06, 0.65)	0.581	0.012	0.28 (0.07, 0.79)	0.591	0.029
Age						
age1	2.07 (1.19, 4.58)	0.336	0.031	1.87 (1.09, 4.08)	0.328	0.056

3.2.3.2 Model Validation

3.2.3.2.1 Assessing the Homoscedasticity Assumption

The spread of the residuals in Figure 73 seems acceptable for our binary logistic regression models. Therefore, any obvious violation of the homoscedasticity assumption was ruled out. For the rest of the analysis, we will be focusing on the fixed effects model that was fitted to the entire dataset.

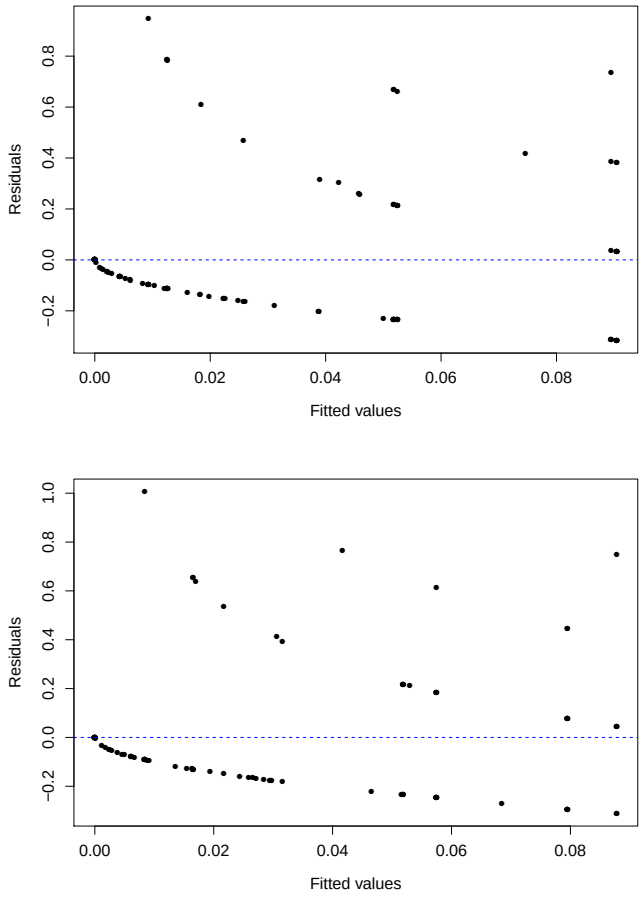


Figure 73: Grouped Pearson residuals versus fitted values for the model that was fitted to the entire dataset (top) and the model that was fitted to unique observations per household (bottom).

3.2.3.2.2 Assessing the Assumption of Independence

Figure 74 suggests a segregation of negative and positive residuals in the urban and peri-urban sites. Compared to Figure 75, our provisional model mainly over-estimated the risk of *S. haematobium* infection in the urban and peri-urban sites. The rural sites, however, did not show any clear clustering of risk, except for one community where the likelihood of infection seemed to have been under-estimated by the provisional model.

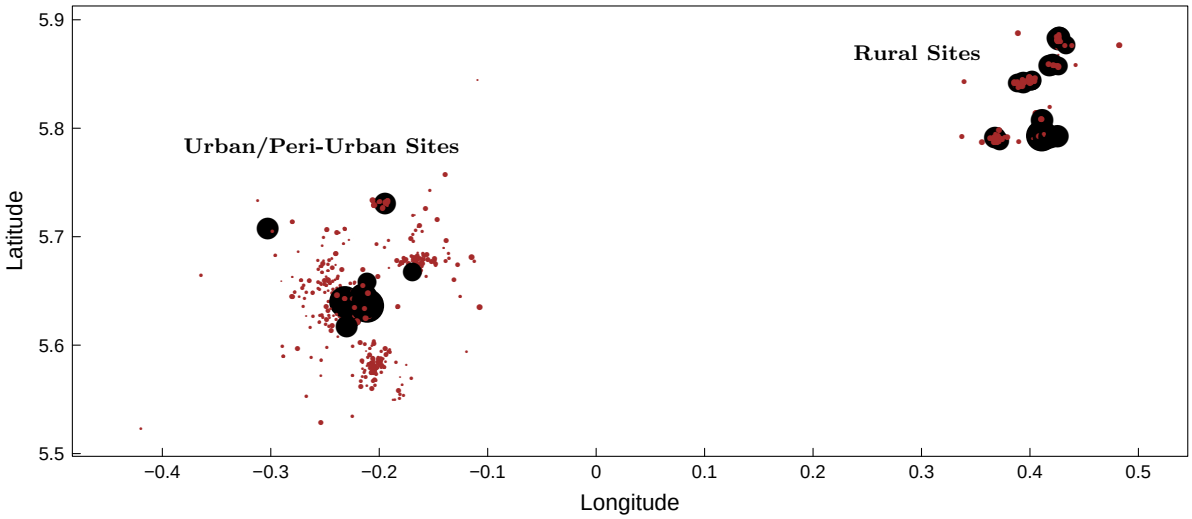


Figure 74: Spatially-referenced residuals of the provisional model for *S. haematobium* mono-infections in the three sites. The radii of the circles are proportional to the absolute values of the Pearson residuals. The red and black circles represent the negative and positive residuals, respectively.

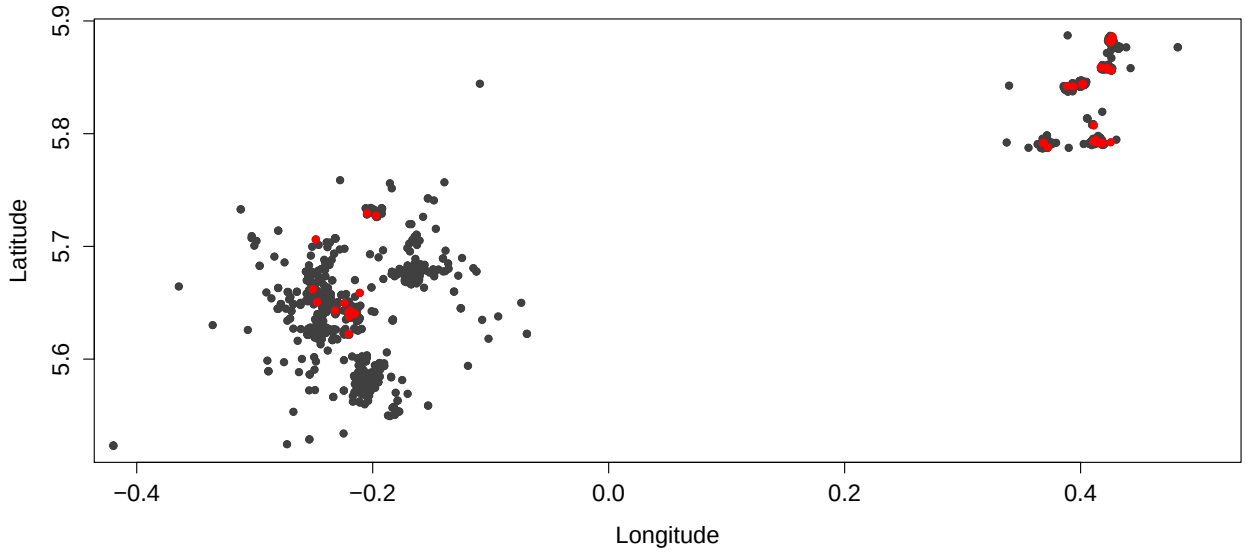


Figure 75: Distribution of *S. haematobium* mono-infections in the three sites. Each point corresponds to a sampled location i.e location of the household of the sampled child. The red and black points represent infected and non-infected cases, respectively.

To help provide a better and more objective assessment of the spatial dependence, an empirical variogram was computed.

3.2.3.2.3 Structural Analysis

The omni-directional empirical variogram, (Figure 76 (A)), indicates a weak spatial trend where the nugget effect accounted for about 90 percent of the total variation in the residuals. Our assessment of directional dependence suggests the isotropy assumption seemed reasonable (Figure 76 (B)). Therefore, the omni-directional empirical variogram was regarded as a mean variogram for all spatial directions and used in the rest of the analysis. Both the envelope of random permutations, Figure 77, and our formal test for trend by Eagle & Diggle 2012 confirmed the presence of a marginally significant spatial trend (p=0.093).

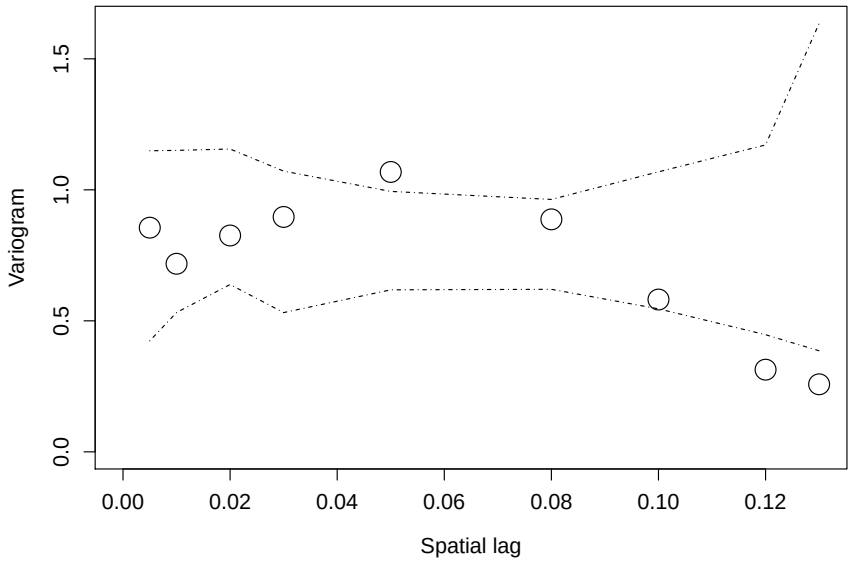


Figure 77: Empirical variogram of the Pearson residuals of the provisional model for *S. haematobium* mono-infections, ∘ ∘ , and 95% confidence envelope. The shape variogram suggests the presence of a marginal spatial trend.

3.2.3.2.4 Estimated Covariance Parameters and Final Model

The parameters of the covariance structure were estimated from the fitted parametric model, Figure 78, as $\phi = 0.006$, $\sigma^2 = 0.078$ and $\tau^2 = 0.729$; by weighted least squares. Table 51 presents a comparison of the final geostatistical and the provisional models. Due to the marginal spatial correlation, the predicted point estimates of *S. haematobium* infection, Figure 79, were similar to the observed infections.

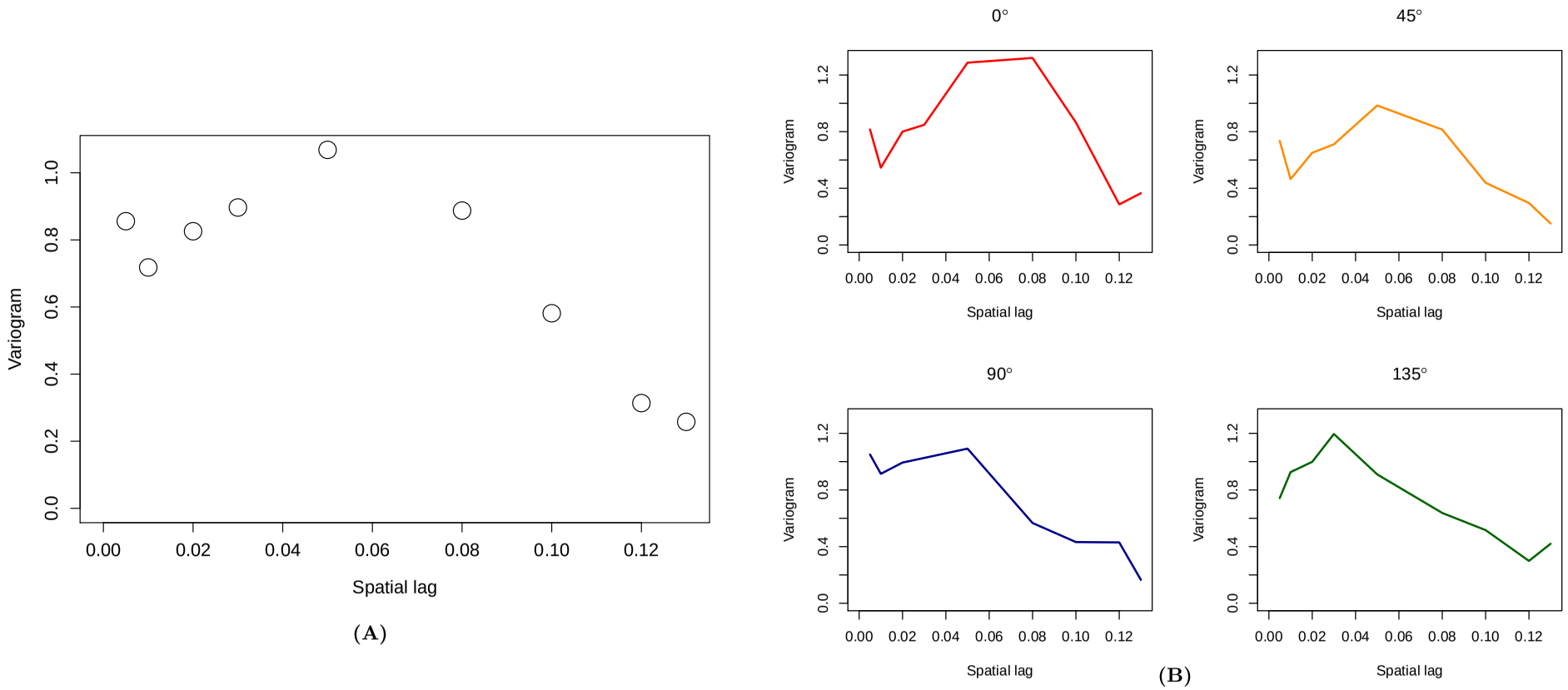


Figure 76: (A) Omni-directional empirical variogram for the Pearson residuals of the provisional model for *S. haematobium* mono-infections. (B) Directional empirical variograms for lags in the 0°, 45°, 90° and 135° spatial directions. The directional variograms are relatively similar to the isotropic variogram. Therefore, the isotropy assumption seems reasonable.

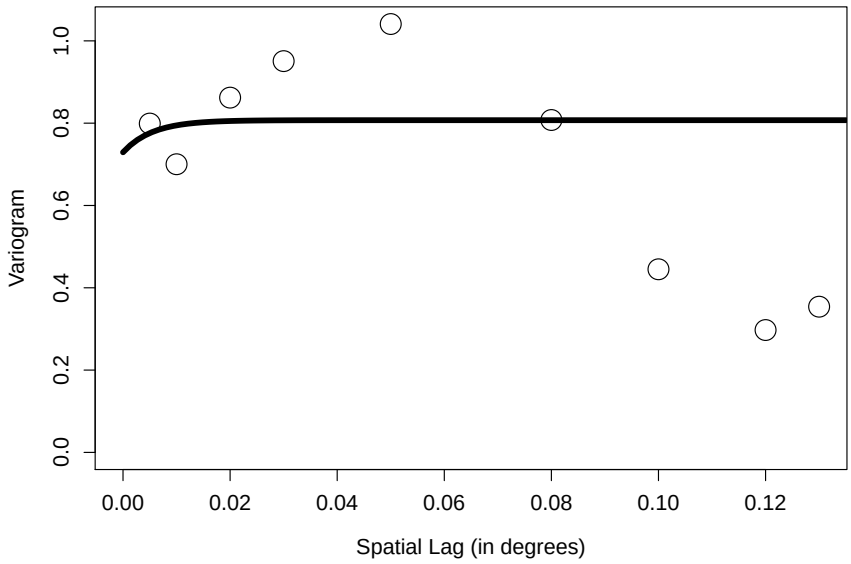


Figure 78: Empirical variogram for *S. haematobium* mono-infections with the fitted Matern covariance function of the order $\kappa = 0.5$. The estimated covariance parameters are: $\phi = 0.006$, $\sigma^2 = 0.078$, $\tau^2 = 0.729$.

Table 51: Regression parameters and standard errors of the final and provisional models for *S. haematobium* mono-infections.

	Geostatistical Model			**Provisional Model**		
Parameters	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
No formal education	1					
Basic	-0.557	0.346	0.107	-0.587	0.301	0.052
Secondary	-0.883	0.675	0.191	-0.890	0.596	0.135
Further	-1.681	1.063	0.114	-1.666	0.950	0.080
Higher	-15.874	1310.051	0.990	-15.874	722.821	0.983
Charcoal	1					
Firewood	0.052	0.348	0.881	0.012	0.301	0.967
Liquified petroleum gas	-1.486	0.654	0.023	-1.460	0.581	0.012
age1	0.756	0.380	0.047	0.727	0.336	0.031
Covariance Parameters						
log(sigma2)	-2.064	0.487				
log(phi)	-3.715	0.602				

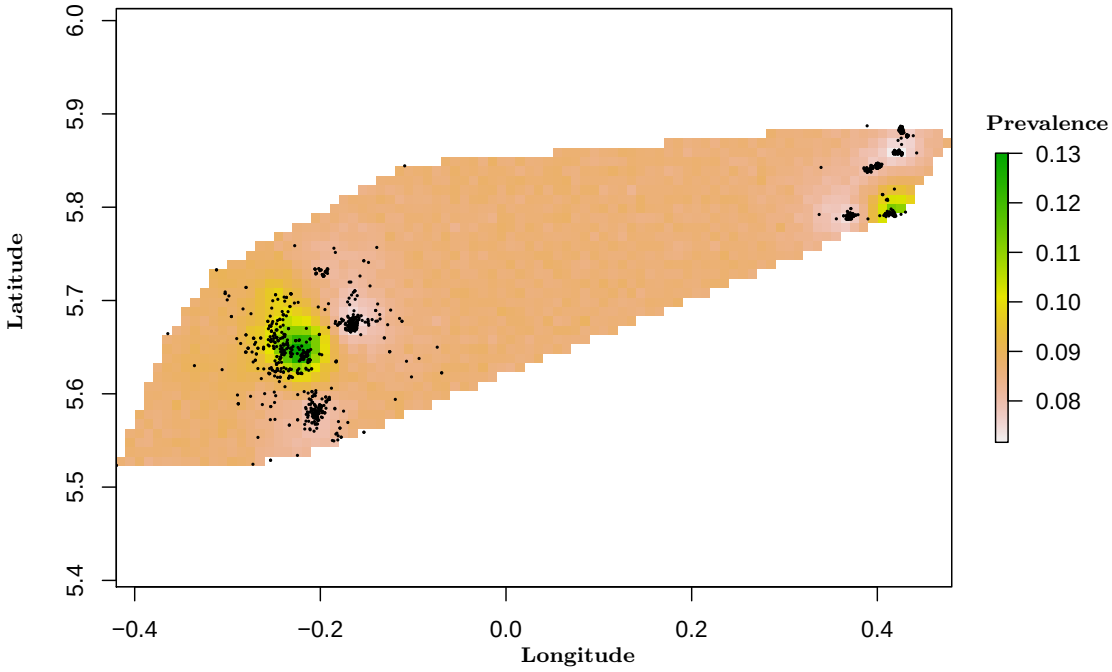


Figure 79: Predicted prevalence of *S. haematobium* infections based on the geostatistical binomial model (Table 51). The black dots represent the surveyed areas.

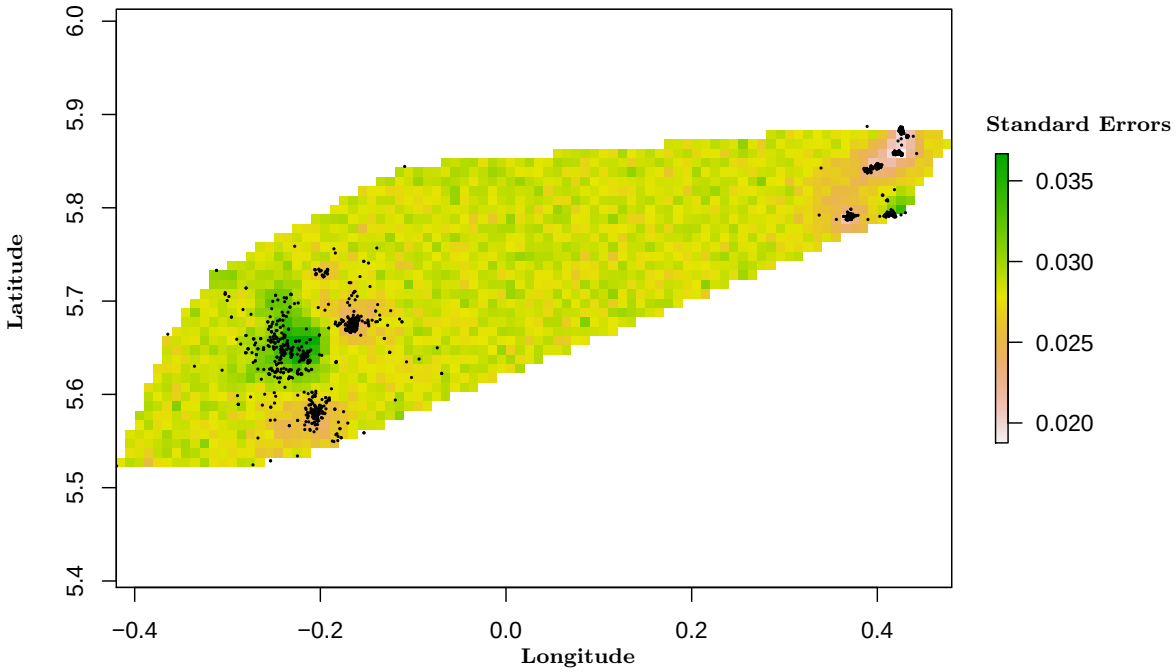


Figure 80: Standard errors associated with the predicted prevalence of *S. haematobium* infections.

4 Discussion

Renewed interest in the control of schistosomiasis and the soil-transmitted helminth infections, following the endorsement of the World Health Assembly's resolution 54.19 by Member States, led to the formation of the Partners for Parasite Control and the Integrated Control of Schistosomiasis in sub-Saharan Africa (WHO 2005, SCI 2014). These organisations have, therefore, been responsible for most of the large scale chemotherapy intervention campaigns across the endemic parts of sub-Saharan Africa since the mid-2000's. Moreover, since school-aged children, 6-16 years, constitute the most susceptible human hosts population for schistosomiasis and the STH infections, these organisations have mainly targeted their key interventions at children at the basic school level.

However, while these concerted intervention campaigns may have succeeded in reducing the disease burden in endemic areas, control programmes have been faced with the much greater challenge of consolidating their achievements in order to avoid the risk of resurgence. This is mainly due to the fact that despite the declining transmission rates in these post-intervention areas, the public health threat posed by schistosomiasis cannot be completely overruled. But rather, transmissions may continue to occur in these areas at a decreased rate which may eventually lead to the resurgence of infections up to the scale of their pre-intervention levels (WHO 2002). Therefore, instead of relaxing the level of vigilance in post-intervention areas, more robust strategies may be required for identifying and directing interventions at the sparsely distributed remaining reservoirs of transmission before the focus of control programmes completely shifts to other pre-intervention areas.

The present study has, therefore, investigated the role of socio-economic factors in predicting the distribution of the reservoirs of schistosome and the soil-transmitted helminths transmissions in three districts with low endemicity levels in the Greater Accra Region of Ghana. Our study sites were chosen to reflect the different conditions typical to low schistosome endemicity settings. Therefore, whereas the low transmission rates in the rural sites were mainly due to previous school-based chemotherapy intervention campaigns, transmission in the peri-urban site could be attributed to both autochthonous transmission and infections among recent migrants from rural areas. Such low trans-

mission foci in peri-urban areas have also been reported by studies in other endemic parts of sub-Saharan Africa (Bockarie et al. 2013, Matthys et al. 2007, Mott et al. 1990).

Despite these established transmission foci in peri-urban areas, however, the rural parts of endemic countries have mainly been the focus of chemotherapy interventions due to their high transmission rates. Hence, infected cases in urban and peri-urban areas are often left to seek their own treatment. Therefore, the range of measured socio-economic factors in this study were meant to serve as proxies for the conditions of the households in the surveyed areas; and how these conditions may have influenced access to improved sanitation, potable water, level of health awareness and health-seeking behaviours.

For instance, the level of health awareness at the household level may influence the infection status of any given child. This is mainly due to the fact that some households may follow regular treatment regimes with over-the-counter medication based on their level of health education. Moreover, as part of their domestic chores, children in rural communities are responsible for collecting water for their households. Therefore, their risk of schistosome infection automatically increases when the water has to be collected from surface water sources. Hence, restricted access to improved water and sanitation results in heterogeneities in schistosome transmission.

Our study objectives were achieved to the following extent: whereas evidence of residual spatial variation was found in the transmission of the soil-transmitted helminth (STH) infections, the transmission of *S. haematobium* infections showed only marginally significant residual spatial variation. Therefore, the predicted prevalence of *S. haematobium* infections may be comparable to the observed infections based on the data.

Though the risk map for the STH infections mainly predicted transmission rates below 3.2%, these predictions could only be interpreted in the context of the quality of the data they are based on. For instance, the duration between slide preparation and the reading of the slides in the laboratory could influence which of the STH infections are detected. Moreover, judging by the fact that the studied areas were low transmission sites, a much more sensitive diagnostic technique, other than microscopy, would have fared a lot better in providing more accurate estimates of the transmission rates for

both *S. haematobium* and the STH infections. Therefore, the predicted transmission rates for the STH infections may be associated with some amount of variability.

By using a set of covariates that influence the persistence of infections across different settings, we may have succeeded in transcending the setting-specific determinants of infectivity in endemic areas; provided the measured covariates actually reflected the socio-economic conditions of households. Moreover, the model-based approach of predicting infections enables the heterogeneities in the transmission of the studied infections to be taken into account, thereby producing more accurate predictions. However, the use of microscopy as our diagnostic tool could be regarded as a possible limitation which may have influence the outcome of this study. For instance, the transmission rate of *S. haematobium* infections may be much higher in the studied areas.

4.1 Conclusions and Recommendations

We have predicted the transmission pattens of schistosomes and the soil-transmitted helminth infections in low endemicity areas. These predictions were based on models that assumed that persisting infections in low transmissions areas were influenced by the socio-economic standards of households that determine assess to sanitation and potable water. However, though this assumption could be regarded as a way of transcending the setting-specific variations in the factors that influence schistosome transmissions, it may have some limitations.

For instance, perennial shortage of water and the rationing of potable water is common in the urban and peri-urban areas of endemic countries. Therefore, people from different socio-economic backgrounds may resort to using untreated or inadequate quantities water at different points in time. Hence, poor hygiene practices, resulting from inadequate supply of water, may occasionally apply across different households irrespective of their socio-economic standards.

Future studies in this area would benefit from longitudinal data that capture information such as assess to water in the urban and peri-urban areas at different time points and seasons. Moreover, more sensitive diagnostic techniques for the detection of light intensity infections in such areas would help improve the accurately of the predictions.

References

Arlinghaus, S. L., ed. (1996), *Practical Handbook of Spatial Statistics* , CRC Press.

Basanez, M.-G., French, M. D., Walker, M. & Churcher, T. S. (2012), 'Paradigm lost: how parasite control may alter pattern and process in human helminthiases', *Trends in Parasitology* **28**(4), 161 – 171.

Bates, D., Maechler, M. & Bolker, B. (2012), *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999999-0.

Bockarie, M., Kelly-Hope, L., Rebollo, M. & Molyneux, D. (2013), 'Preventive chemotherapy as a strategy for elimination of neglected tropical parasitic diseases: endgame challenges', *Philos Trans R Soc Lond B Biol Sci* **368**(1623).

Brooker, S. (2007), 'Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control', *Transactions of The Royal Society of Tropical Medicine and Hygiene* **101**(1), 1–8.

Brooker, S., Hay, S. I. & Bundy, D. A. (2002), 'Tools from ecology: useful for evaluating infection risk models?', *Trends in Parasitology* **18**(2), 70 – 74.

Brooker, S., Kabaterein, N. B., Gyapong, J. O., Stothard, J. R. & Utzinger, J. (2009), 'Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa', *Parasitology* **136**, 1707–1718.

Bruun, B. & Aagaard-Hansen, J. (2008), *The social context of schistosomiasis and its control: An introduction and annotated bibliography*, World Health Organization.

Clements, A., Bosque-Oliva, E., Sacko, M., Landoure, A., Dembele, R., Traore, M., Coulibaly, G., Gabrielli, A., Fenwick, A. & Brooker, S. (2009), 'A comparative study of the spatial distribution of schistosomiasis in Mali in 1984-1989 and 2004-2006', *Plos Neglected Tropical Diseases* **3**(5), e431.

Clements, A. C. A., Lwambo, N. J. S., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A. & Brooker, S. (2006), 'Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis

control programme in Tanzania', *Tropical Medicine & International Health* **11**(4), 490–503.

Collett, D. (2003), *Modeling Binary Data*, Texts in Statistical Science Series, Chapman and Hall/CRC, 1ccop. 2003.

Cowper, S. (1971), *A Synopsis of African Bilharziasis*, H. K. Lewis.

Crawley, M. (2002), *Statistical Computing: An Introduction to Data Analysis using S-Plus*, Wiley.

CRED (2013), 'Ghana - Dara', `http://daraint.org/wp-content/uploads/2013/12/rri-ghana.pdf`.

Diggle, P. J. & Ribeiro, P. J. (2007), *Model-based Geostatistics*, Springer.

Eagle, S. & Diggle, P. J. (2012), Spatial independence test for Geo-statistical data, Technical report, University of Liverpool.

Engels, D., Chitsulo, L., Montresor, A. & Savioli, L. (2002), 'The global epidemiological situation of schistosomiasis and new approaches to control and research', *Acta Tropica* **82**(2), 139 – 146.

Filmer, D. & Pritchett, L. (2001), 'Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India', *Demography* **38**(1), 115–132.

Gazzinelli, A., Correa-Oliveira, R., Yang, G., Boatin, B. & Kloos, H. (2012), 'A research agenda for helminth diseases of humans: social ecology, environmental determinants, and health systems.', *PLoS Neglected Tropical Diseases* **6**(4), e1603.

Giorgi, E. & Diggle, P. J. (2014), *PrevMap: - an R package for prevalence mapping*. R package version 1.0.

GLOFAL (2006), 'Global View of Food Allergy', `http://www.glofal.org/about.html`.

Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006), 'Human schistosomiasis', *The Lancet* **368**(9541), 1106 – 1118.

GSS (2012), '2010 Population and housing census: summary report of final results', http://www.statsghana.gov.gh/docfiles/2010phc/Census2010_Summary_report_of_final_results.pdf.

Hilbe, J. (2009), *Logistic Regression Models*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Hogewoning, A., Larbi, I., Addo, H., Amoah, A., Boakye, D., Hartgers, F., Yazdanbakhsh, M., Van Ree, R., Bouwes Bavinck, J. & Lavrijsen, A. (2010), 'Allergic characteristics of urban schoolchildren with atopic eczema in Ghana', *Journal of the European Academy of Dermatology and Venereology* **24**(12), 1406–1412.

Huang, Y. & Manderson, L. (1992), 'Schistosomiasis and the social patterning of infection', *Acta Tropica* **51**, 175 – 194.

JHR (n.d.), 'Islamic Schools in Ghana - Educating a Minority', http://www.jhr.ca/blog/2011/07/islamic-schools-in-ghana-educating-a-minority/.

Jordan, P. & Webbe, G. (1982), *Schistosomiasis: epidemiology, treatment, and control*, Heinemann.

Kahle, D. & Wickham, H. (2013), *ggmap: A package for spatial visualization with Google Maps and OpenStreetMap*. R package version 2.3.

Kloos, H. (1995), 'Human behavior, health education and schistosomiasis control: A review', *Social Science & Medicine* **40**(11), 1497 – 1511.

Kreft, I. G. & de Leeuw, J. (1998), *Introducing Multilevel Modeling*, SAGE Publications Ltd.

Lammie, P. J., Fenwick, A. & Utzinger, J. (2006), 'A blueprint for success: integration of neglected tropical disease control programmes', *Trends in Parasitology* **22**(7), 313 – 321.

Leyland, A. & Goldstein, H., eds (2001), *Multilevel Modelling of Health Statistics*, Wiley series in probability and statistics, Wiley, Chichester, UK.

Lustigman, S., Prichard, R., Gazzinelli, A., Grant, W., Boatin, B., McCarthy, J. &

Basanez, M. (2012), 'A research agenda for helminth diseases of humans: the problem of helminthiases.', *PLoS Neglected Tropical Diseases* **6**(4), e1582.

Mahmoud, A. (2001), *Schistosomiasis*, Tropical Medicine: Science and Practice, Imperial College Press.

Map Library (2014), 'Ghana - Greater Accra', `http://www.maplibrary.org`.

Matthys, B., Tschannen, A. B., Tian-Bi, N. T., Comoe, H., Diabate, S., Traore, M., Vounatsou, P., Raso, G., Gosoniu, L., Tanner, M., Cisse, G., N'Goran, E. K. & Utzinger, J. (2007), 'Risk factors for *Schistosoma mansoni* and hookworm in urban farming communities in western Côte d'Ivoire', *Tropical Medicine & International Health* **12**(6), 709–723.

Mott, K. E., Desjeux, P., Moncayo, A., Ranque, P. & de Raadt, P. (1990), 'Parasitic diseases and urban development', *Bulletin of the World Health Organization* **68**(6), 691–698.

Pebesma, E., Cornford, D., Dubois, G., Heuvelink, G., Hristopoulos, D., Pilz, Jand Stoehlker, U., Morin, G. & Skoien, J. (2010), 'Intamap: the design and implementation of an interoperable automated interpolation web service', *Computers & Geosciences* **37**, 343–352.

Pinheiro, J. & Bates, D. (2000), *Mixed-Effects Models in S and S-PLUS*, Statistics and computing, U.S. Government Printing Office.

Ribeiro Jr, P. J. & Diggle, P. J. (2001), 'geoR: a package for geostatistical analysis', *R-NEWS* **1**(2), 14–18. ISSN 1609-3631.

Rutstein, S. O. (2008), 'The DHS Wealth Index: Approaches for Rural and Urban Areas', `http://pdf.usaid.gov/pdf_docs/PNADN521.pdf`.

Rutstein, S. O. & Johnson, K. (2004), 'The DHS Wealth Index. DHS Comparative Reports No. 6 ', `http://www.childinfo.org/files/DHS_Wealth_Index_(DHS_Comparative_Reports).pdf`.

SCI (2014), 'ICOSA', `https://www3.imperial.ac.uk/schisto/wherewework/dfid`.

Self, S. & Liang, K.-Y. (1987), 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions', *Journal of the American Statistical Association* **82**(398), 605–610.

Tchuem Tchuente, L.-A., Momo, S. C., Stothard, J. R. & Rollinson, D. (2013), 'Efficacy of praziquantel and reinfection patterns in single and mixed infection foci for intestinal and urogenital schistosomiasis in Cameroon', *Acta Tropica* **128**(2), 275 – 283.

Thangarajan, M. (2007), *Groundwater: Resource Evaluation, Augmentation, Contamination, Restoration, Modeling and Management*, Springer Netherlands.

UNICEF (2006), 'Progress for Children: A Report Card on Water and Sanitation', `http://www.unicef.org/publications/files/Progress_for_Children_No._5_-_English.pdf`.

Utzinger, J., Bergquist, R., Shu-Hua, X., Singer, B. H. & Tanner, M. (2003), 'Sustainable schistosomiasis control − the way forward', *The Lancet* **362**(9399), 1932 – 1934.

Waller, L. A. & Gotway, C. A. (2004), *Spatial Exposure Data*, John Wiley & Sons, Inc., chapter 8, pp. 272–324.

Webster, R. & Oliver, M. (2001), *Geostatistics for environmental scientists*, Statistics in practice, John Wiley & Sons.

White, P., Webber, D. J. & Helvin, A. (2014), *Modellling structural change using broken sticks*, University of the West of England, Bristol, UK.

WHO (1985), 'The Control of Schistosomiasis', `http://whqlibdoc.who.int/trs/WHO_TRS_728.pdf`. Accessed: 12/04/2012.

WHO (1987), 'Atlas of the global distribution of schistosomiasis: 15 - GHANA', `http://whqlibdoc.who.int/temp/schistosomiases/2867810604_15.pdf`.

WHO (2002), 'Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of a WHO expert committee. WHO Technical Report Series No. 912', `http://whqlibdoc.who.int/trs/WHO_TRS_912.pdf`.

WHO (2005), 'Report of the Scientific Working Group Meeting on Schistosomiasis: 14-16 November 2005, Geneva, Switzerland', http://www.who.int/tdr/publications/documents/swg_schisto.pdf.

WHO (2013*a*), Report of an informal consultation on schistosomiasis control, Technical report, World Health Organization.

WHO (2013*b*), Schistosomiasis: Progress report 2001-2011 and strategic plan 2012-2020, Technical report, Geneva, Switzerland.

Wikipedia (2014*a*), 'Songor Lagoon', http://en.wikipedia.org/wiki/Songhor_Lagoon. [Online; accessed 15-October-2014].

Wikipedia (2014*b*), 'Tro-tro', http://en.wikipedia.org/wiki/Tro_tro. [Online; accessed 21-April-2014].

Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Woolhouse, M. E. J., Etard, J.-F., Dietz, K., Ndhlovu, P. D. & Chandiwana, S. K. (1998), 'Heterogeneities in schistosome transmission dynamics and control', *Parasitology* **117**, 475–482.

Zuur, A., Hilbe, J. & Ieno, E. (2013), *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists*, A Beginner's Guide to, Highland Statistics.

Zuur, A., Ieno, E. & Smith, G. (2007), *Analyzing Ecological Data*, Statistics for Biology and Health, Springer New York.

Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. (2011), *Mixed Effects Models and Extensions in Ecology with R.*, Springer.

Part III

Spatial heterogeneity in the infectivity of water contact sites and the risk of harbouring aggregated schistosome burden

Abstract

Background Human exposure activities in endemic shoreline settings are exclusively confined to specially reserved parts of the littoral zone known as water contact sites. The littoral zone also favours the survival of the intermediate snail hosts and the free-living larval stages of the schistosome species. Therefore, both human-to-snail and snail-to-human transmission of infections occur in these water contact sites. Different sites also tend to be reserved for specific purposes. Hence, it follows that if the different water contact sites within any defined endemic setting were regarded as discrete entities then their levels of infectivity would vary to reflect the ecological changes induced by the different exposure activities on the snail host species.

Methods Using a pre-intervention shoreline community of the Volta Lake as our study site, the risk of aggregated *S. haematobium* burden was modelled on individual-level covariates measured during the period of stagnation of the Lake. Factors that influenced the infectivity of the water contact sites were, however, regarded as unobserved factors. Given the exposure dynamics, where members the same familial units mainly frequented common water contact sites, we logically assumed the likelihood in risk would vary across different households based on the infectivity of the point of exposure. Therefore, the ecological effect of exposure activities could be assessed as latent patterns in the points-referenced residuals.

Results The seven water contact sites were associated with varying odds of decreased risk. Moreover, being in the age group of 5-12 years increased the odds by 1.34 (95% CI: 1.11, 1.61) whilst males had a 2.30 (95% CI: 1.21-4.36) greater odds of harbouring aggregated burden. There was, however, no statistically significant latent patterns in the residuals (p=0.888). Our results, therefore, seem to suggest the mean burden of infection varied within familial units, despite their common points of exposure.

Conclusions The burden of infection within the human host population did not seem to have served as a good marker for the infectivity of water contact sites. This study would have, however, benefited from a much better assessment of how diffused or concentrated exposure patterns were and multiple surveys at different time points during the period of stagnation. We, therefore, consider our current findings inconclusive.

1 Introduction

The parasitic schistosomes of humans are known to be resilient to both natural and anthropogenic perturbations, including those induced by chemotherapy interventions. This resilience is attributed to their highly regulated distribution within the human and snail host populations (Anderson & Medley 1985, Lustigman et al. 2012). Consequently, about 80% of the schistosome burden within any endemic setting is estimated to be harboured by just 15 - 20% of the infected human host population (Jamison et al. 2006, Anderson & Medley 1985). However, while this aggregation of heavy intensity infections may pose challenges to control when such individuals are missed during chemotherapy intervention programmes, the direct targeting of interventions at these heavily infected individuals could also maximise the effectiveness of control programmes (Gurarie & King 2005).

It is well-established that the age-intensity profiles of infections also tend to be convex-shaped, with children harbouring most of these aggregated schistosomes burdens while light intensity infections tend to be more common among the older age-groups (Anderson & Medley 1985). Therefore, control programmes have utilised both school-based and community-wide intervention administration strategies, with children as the primary focus of control (Gurarie & King 2005, SCI 2013).

What remains poorly understood though are the processes that underlie the predisposition to aggregated schistosome burdens within the human host population in different endemic settings (Jamison et al. 2006). The fact that the re-infection rates, following chemotherapy interventions, tend to be slower among the older age-groups irrespective of their exposure patterns; has led to suggestions that genetic variability, resulting mainly from acquired immunity, may underlie these observed predispositions to aggregated schistosome burdens (Jamison et al. 2006, Anderson & Medley 1985). However, while the role of acquired immunity may point to the selective targeting of interventions at heavily infected cases as the most effective means of control, there have been suggestions that community-wide intervention campaigns in endemic areas could also reverse the effect of the apparent "herd immunity" among the older age-groups. This induced susceptibility is generally thought to be the direct consequence of repeating

chemotherapy interventions at frequencies that are below the required threshold level for the elimination of infections within any defined endemic setting (Anderson & Medley 1985).

Therefore, even though the frequency of administrating chemotherapy interventions within different endemic settings is established using endemicity levels, prediction of the efficacy of these interventions is rendered virtually impossible by the combined effect of the aggregated schistosome burdens and apparent "herd immunity" among the infected human host population (Lustigman et al. 2012, Anderson & Medley 1985). Consequently, ongoing comparison trials of school-based and community-wide strategies of administering chemotherapy interventions, by the Schistosomiasis Consortium for Operational Research and Evaluation (SCORE), are mostly necessitated by some of such complications in the prediction of the outcomes of interventions (SCI 2013).

However, although acquired immunity may play a role in explaining the predisposition to aggregated schistosome burdens, transmissions are solely dependent on exposure to cercaria-infested freshwater bodies. Therefore, the varying degree and pattern of aggregation of these parasites within the human host population in different endemic settings may lend credence to the role of varying exposure patterns as well as acquired immunity (Guyatt et al. 1994, Chandiwana & Woolhouse 1991). Indeed, Bruun & Aagaard-Hansen 2008 contended, after an extensive review of water contact studies, that varying exposure patterns may be the most important determinants of the predisposition to aggregated schistosome burdens.

Compelling arguments have also been offered in favour of the fact that human exposure patterns in endemic areas may principally be age dependent, with the maximum rates of exposure occurring among older children and young adults (Dalton & Pole 1978, Scott et al. 1982, Anderson & Medley 1985), thereby lending credence to the convex-shaped age-intensity profiles. However, while this argument may prove true for most endemic settings, it is also an established fact that infectivity is not guaranteed by every water contact activity (Bruun & Aagaard-Hansen 2008). But rather, infectivity may be influenced by an interaction between a series of favourable environmental, biological and social processes that determine the convergence of the human hosts with the infective

stages of the schistosome parasites as well as the snail hosts in space and time (Bruun & Aagaard-Hansen 2008, Kloos et al. 1998). In this regard, exposure is viewed in the context of activities that involve both human-to-snail as well as snail-to-human transmission of infections (Figure 81). Therefore, even though factors such as the frequency, duration, time and degree of exposure to suitable water bodies have often been associated with heavy intensity infections (Chandiwana & Woolhouse 1991, Dalton & Pole 1978, Kloos et al. 1998), the likelihood of harbouring aggregated schistosome burden as a consequence of repeated exposure would depend largely on whether all the prerequisites for infection are met on every exposure event. However, the subject of whether these prerequisites could be met on every exposure is open to debate.

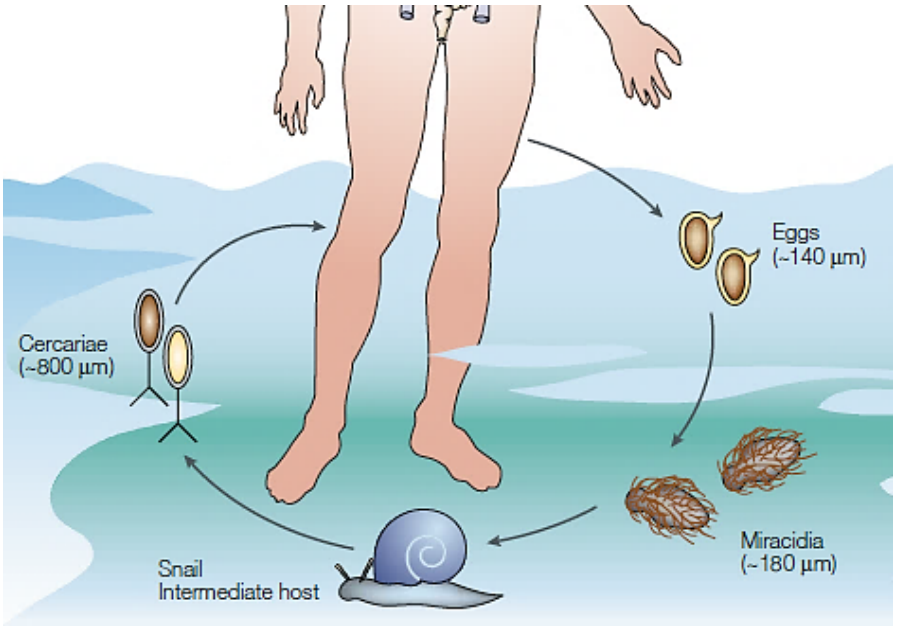


Figure 81: Life cycle of the schistosome parasites showing the role of the intermediate snail hosts in the transmission of infections (adapted from Pearce & MacDonald 2002 with permission of the rights holder, Nature Publishing Group).

Given their crucial role in the life cycle of the human schistosome parasites, the presence of the intermediate snail hosts within any endemic setting tends to signify the transmission of infections (Figure 81). It follows that any conditions that influence the ecology and bionomics of these snails in their freshwater habitats would also have important implications on the levels of infectivity of these water bodies as well as the overall transmission rate of *Schistosoma* infections in any endemic setting (Cowper 1971). While the survival of these snails may be influenced by a myriad of physical, chemical and biolo-

gical factors, they are mainly known to thrive under aquatic conditions of low flow rates of less than 15 metres per minute. Hence, large colonies of these snails tend to establish themselves in the littoral zones of their freshwater habitats (Cowper 1971).

However, the littoral zones of lakes in endemic settings also serve as water contact sites, which are specially reserved areas for a series of domestic, recreational and economic activities by the inhabitants of shoreline communities (Cowper 1971, Dalton & Pole 1978). Therefore, all exposure activities involving both human-to-snail as well as snail-to-human transmission of infections in such areas are exclusively confined to these water contact sites (Chu & Vanderburg 1976). Moreover, studies in different endemic settings have consistently shown that different water contact sites may be reserved for different purposes by the inhabitants of such settings (Bruun & Aagaard-Hansen 2008). Hence, significant variations may be observed in the amount of exposure activities that are performed by different subsets of the population at different water contact sites (Chandiwana & Woolhouse 1991).

While these water contact sites may be the point of a series of setting-specific activities, the effect of these activities on the ecology of the snail hosts that inhabit the littoral zones, may differ considerably. For instance, whereas the pollution of these water contact sites by human excreta, which mainly occurs during swimming by children, may favour the establishment of the snail hosts; anthropogenic changes resulting from chemical pollution due to washing or bathing with soap as well as the artificial turbulence that is constantly created by the passage of boats and wading may adversely affect the survival of these snails (Cowper 1971).

Moreover, even though the net rate of contamination with human excreta may differ for different water contact sites, the susceptibility of the snail hosts in the various sites to infection by the miracidia (Figure 81), may be determined by their natural immunity as well as a series of physical and chemical conditions specific to the water contact sites; among which light intensity and temperature tend to be the most important (Woolhouse et al. 1998, Cowper 1971). Therefore, different water contact sites within the same endemic setting may exhibit strong heterogeneities in both their snail and temporal cercarial densities, and consequently vary in their levels of infectivity (Gryseels et al. 2006).

Therefore, if the different water contact sites within any defined endemic shoreline setting were regarded as discrete entities with varying levels of infectivity, resulting from the cumulative effect of anthropogenic activities as well as the physical and biological conditions unique to these water contact sites, then the different subsets of the human hosts population that frequent these sites would also experience varying degrees of infections. Hence, despite the generally recognised age-specific exposure patterns and age-specific susceptibility within the human hosts population, the risk of harbouring aggregated schistosome burdens may also be a function of where exposure activities are performed.

However, since factors that influence the varying infectivity of different water contact sites cannot be measured quantitatively with ease, most studies that have investigated the risk of aggregated schistosome burdens have mainly relied on the more common exposures such as age, frequency, duration, and degree of water contact (Chandiwana & Woolhouse 1991, Dalton & Pole 1978, Kloos et al. 1998). Moreover, even though the role of genetic variability in the immunological competence of the human hosts population is a widely recognised risk factor, its effect is often assumed to be negligible in models (Anderson & Medley 1985), thereby making age-related changes in exposure patterns the most common explanation for the observed convex-shaped age-intensity profiles in endemic areas.

But how well do models that are only based on the commonly measured exposures fare in providing valuable insights into the complexities of the risk of aggregated schistosome burdens in endemic areas? If such models were limited in their ability to effectively account for the risk of aggregated worm burdens, then there could be implications on the efficacy of intervention programmes whose designs are based on such models (Anderson & Medley 1985). For instance, if the role of genetic variability in immunological competence among the human hosts population is not negligible, as is assumed by most models, then its relative effect with age could potentially have implications on the design of intervention programmes.

Moreover, while human exposure patterns are generally thought to be age-dependent, they may also occur within the much broader context of a series of setting-specific social

processes (Bruun & Aagaard-Hansen 2008). Therefore, if the widely held assumption of decreasing exposure patterns with age were relaxed to accommodate scenarios where exposure patterns were random across all age groups in pre-intervention areas, then investigating the risk of harbouring aggregated schistosome burdens within such settings could provide valuable insights into the underlying factors that govern the predisposition to aggregated schistosome burden, and consequently the driving force behind the transmission dynamics of infections.

Therefore, the primary objective of this paper is to develop a model for the risk of aggregated *Schistosoma haematobium* burden that simultaneously accounts for the effects of the measured covariates as well as relevant unmeasured factors that influence the level of infectivity of the water contact sites where exposure occurs. Our specific objectives are: to model the risk of aggregated *S. haematobium* burden as a function of the measured covariates; to assess the residual spatial effects due to the unmeasured factors as latent patterns in the standardised residuals of the provisional model; and if necessary incorporate the spatially-correlated errors due to the effect of the unmeasured factors. By this approach, we will also be assessing the variation in the risk of aggregated *S. haematobium* burden across our study community. Finally, the implications of our findings will be interpreted in the broad context of schistosomiasis control in sub-Saharan Africa.

2 Materials and Methods

To recap on the main concepts so far, aggregated schistosome burden within the human host population is being regarded as the cumulative effect of repeated exposure to infested freshwater bodies. We are also arguing that the different water contact sites within endemic shoreline settings may exhibit heterogeneities in their level of infectivity due to the effect of the site-specific anthropogenic activities on the ecology of the intermediate snail hosts that inhabit these water contact sites. Therefore, even though human exposure patterns may occur at random, infectivity due to any such encounters may vary depending on where exposure occurs.

2.1 Study Site

Therefore, in testing these hypotheses, we took another look at the community-level survey data for Alabonu, a pre-intervention shoreline community of the Volta Lake in Ghana (Figure 82). Details of the data collection and processing methods have already been presented in Part I, section 2.3 - 2.4, of this thesis. Our study population comprised of the physically-active inhabitants of this community who were within the 5 - 60 years age range.

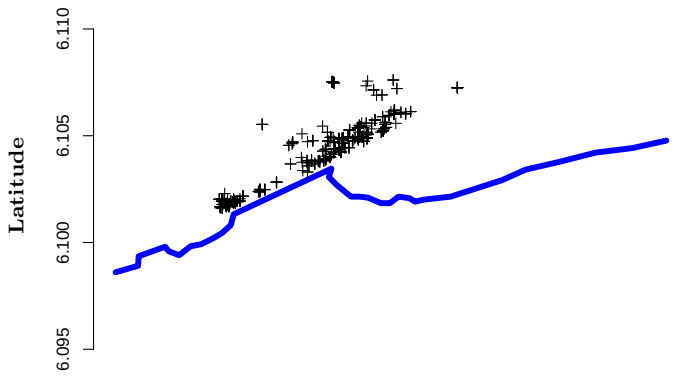


Figure 82: Proximity of households in Alabonu (+) to the main shoreline of the Volta Lake. `R` Packages used in map preparation: `rgdal` (Keitt et al. 2013) and `rgeos` (Bivand & Rundel 2013).

2.1.1 Basis for Selecting Alabonu for this Study

Alabonu was considered suitable for this study for a number of reasons which we now discuss. The pre-intervention status of this community enabled us to assume an endemic equilibrium state for the underlying processes that governed transmissions. Therefore, heterogeneities in the human host population with regard to genetic variability in immunological competence were assumed to be unaffected by any community-level chemotherapy intervention. Additionally, neither the age-intensity profiles nor the density-dependent regulatory mechanisms that influenced the community-level variation in risk had been modified as a consequence of intervention programmes.

(a) (b)

Figure 83: The characteristic stagnant flow rate of the Lake at the water contact sites (Figure 83(a)) and the increased current during the major raining season (Figure 83(b)).

The lack of social amenities such as potable water supply and good access roads made exposure patterns approximately constant among the inhabitants. Moreover, the flow rate of the Lake was virtually stagnant, especially near the margin, for most of the year (Figure 83(a)). This, therefore, fitted into our main assumption of regarding the different water contact sites along the shoreline as discrete entities.

However, the spillways of the Akosombo Dam (Wikipedia 2015), which was situated about six miles upstream, were opened during the major rainy season to release excess water. The resulting change in the hydrology of the Lake potentially had the effect of rendering the site-specific anthropogenic activities negligible and causing fluctuations in

the snail population density (Figure 83(b)). We, however, contend that for at least eight months in each year, the different water contact sites could still be regarded as discrete entities.

2.2 Exposures

Table 52 outlines both the measured and potentially relevant unmeasured covariates that were considered in this study. The measured covariates are based on questionnaire surveys that were conducted alongside the parasitological surveys.

2.3 Study Outcome

Our study outcome, the schistosome burden harboured by the individual study participants, was estimated with the number of excreted *Schistosoma haematobium* eggs in 10 ml urine samples. Though an indirect measure, autopsy studies and more recently, molecular studies have established a direct correlation between egg output and the burden of female *Schistosoma* worms in the body (Cheever et al. 1977). However, there have also been suggestions to the effect that in cases where density-dependent constraints act to regulate the fecundity of the female worms, the egg output may not always be directly proportional to worm burden (Basanez et al. 2012, Anderson & Medley 1985). For the purpose of this study, however, we chose to assume a direct proportionality between egg output and schistosome burden in order to ease the interpretation of our findings.

2.4 Exploratory Analysis

Based on a priori knowledge, we expected a non-linear relationship between age and the discrete values of our study outcome, schistosome burden. This relationship was therefore assessed using non-parametric smoothers, equation 1, as estimated by the `gam` function in the `R mgcv` package (Wood 2006). The estimated smoother subsequently formed the basis for the chosen forms of dependence on age in our formal model.

$$\begin{aligned} Y_i &\sim NB(\mu_i, k) \\ log(\mu_i) &= \alpha + f\left(\mathrm{x}_i\right) \end{aligned} \tag{1}$$

Table 52: The measured and unmeasured exposures that were considered in this study.

Measured Exposures	Description	Unmeasured Exposures
Age	5-60 years	Genetic variability in both the human and snail hosts population
Sex		Pollution of water contact sites due to anthropogenic activities
Frequency of water contact	Daily or weekly contacts at specific water contact sites	Effect of turbulence created by canoes on snail hosts
Water contact site	Specific water contact site frequented	Population density of infected snails per water contact site
Treatment	Treatment with over-the-counter drugs in the past year	Temporal cercarial population density at water contact sites

where the observed outcome in the i^{th} participant, Y_i, was assumed to follow the negative binomial distribution, NB, with expectation, μ_i, and dispersion parameter, k. The intercept of the Generalised Additive Model is denoted by α whilst $f(x_i)$ is the smoothing function for age.

2.5 Confirmatory Analysis

2.5.1 Distribution of Schistosome Burden in the Study Population

An initial assessment of our study outcome indicated that 63% of the participants excreted zero schistosome eggs whilst the egg output in the remaining 37% of the study population ranged between 1 - 8648 (Table 53 and Figure 84). The community-level variance to mean ratio of schistosome burden was approximately 2581.

Table 53: Distribution of worm burden among the 522 participants in this study

Worm Burden	Frequency	Percentage
0	330	63.22
1 - 50	121	23.18
51 - 100	21	4.02
101 - 8648	50	9.58

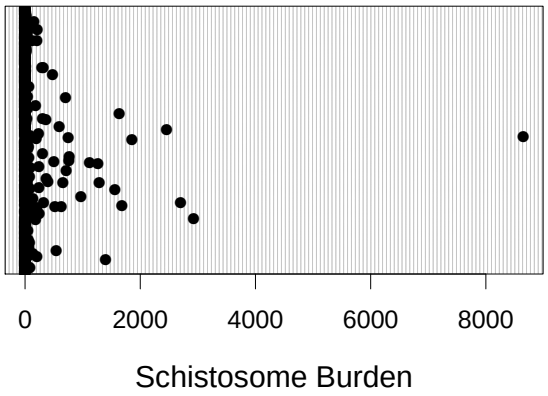


Figure 84: A Cleveland dotplot showing the distribution of worm burden, as measured by *S. haematobium* egg output in 10ml of urine samples per participant.

2.5.1.1 Possible Sources of the Excess Zeros

Given the virtually constant exposure patterns in our study community, it was only normal to initially assume that all the physically-active inhabitants might be infected. However, the majority of our study population turned out to be non-infected (Figure 85). The rest of this section would, therefore, focus on the sub-population of inhabitants who formed the non-responders. We, however, choose at this stage of the analysis to assign negligible significance to the effect of genetic variability in the immunological competence of the human hosts to mount total resistance to infections.

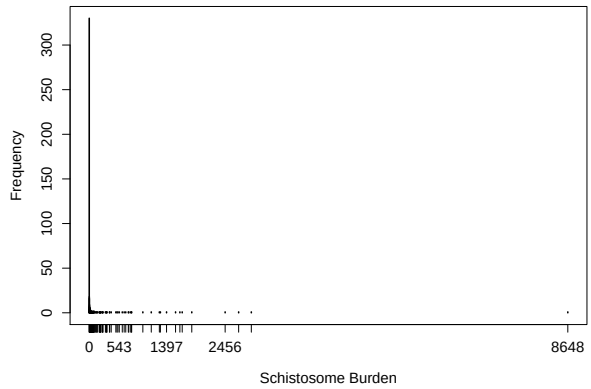


Figure 85: Frequency plot showing the inflated zero counts in the distribution of schistosome burden among the inhabitants of Alabonu.

While the time range for our samples collection, 11:00 - 14:00 GMT, may rule out circadian variation in schistosome egg output among the infected human hosts as a possible source of non-response; the detection method, which was based on the presence of schistosome eggs in urine samples, could potentially have been a contributory factor to the observed non-response rate. Though the presence of eggs in urine samples may be a good criterion for assessing the burden of infections in individuals who harboured paired, adult male and female schistosomes; the criterion may fail to detect unisexual schistosome infections in the human hosts.

Empirical evidence suggests the possibility of harbouring unisexual infections in endemic areas, despite the equal chances of getting infected with both male and female schistosome larvae on every exposure event (Macdonald 1965). Therefore, the sub-population of non-responders in this study may have included individuals who harboured unisexual schistosome infections.

Moreover, judging by the fact that our data came from a high transmission endemic setting, schistosomiasis-associated morbidities such as fibrosis of the bladder were bound to be common among the adult participants. However, this morbidity is known to be associated with a high schistosome egg retention rate; thereby making it another possible source of non-response in this study. Lastly, we also consider ectopic situations where the schistosome eggs get misdirected to organs other than the bladder and hence fail to get excreted in urine (Nasell 1976, Cowper 1971).

2.5.1.2 How Do These Excess Zeros Fit into the Context of this Study?

It would appear from the discussion in section 2.5.1.1 that most of the observed zeros were false zeros, in that zero schistosome eggs were recorded in participants who may have actually harboured infections. However, this study is based on the hypothesis that the various water contact sites in our study community may have exhibited heterogeneities in their levels of infectivity due to the ecological changes induced by the effect of different anthropogenic activities. It is, therefore, logical to assume that the burden of schistosome infections harboured by the individual inhabitants would also vary to reflect the varying infectivity of the specific water contact sites they frequented.

Therefore, the subsets of inhabitants who frequently came into contact with the lake at those sites where anthropogenic activities had an uncongenial effect on the snail host population were most likely to harbour fewer or no infections at all. Consequently, any recorded zero egg counts among those inhabitants may have actually represented true zeros. However, although only one possible source of true zeros is being considered, we also acknowledge the fact that other mechanisms may have contributed to the generation of the true zeros in this study.

2.5.1.3 Model Formulation

Our examination of the possible sources of the observed zero outcomes in section 2.5.1.2 could lead us into regarding these zeros as coming from two main processes: one that generates the false zeros and the other, the true zeros (Zuur et al. 2011). Moreover, the process that generates the true zeros is also responsible for producing the non-zero

observations (Figure 86).

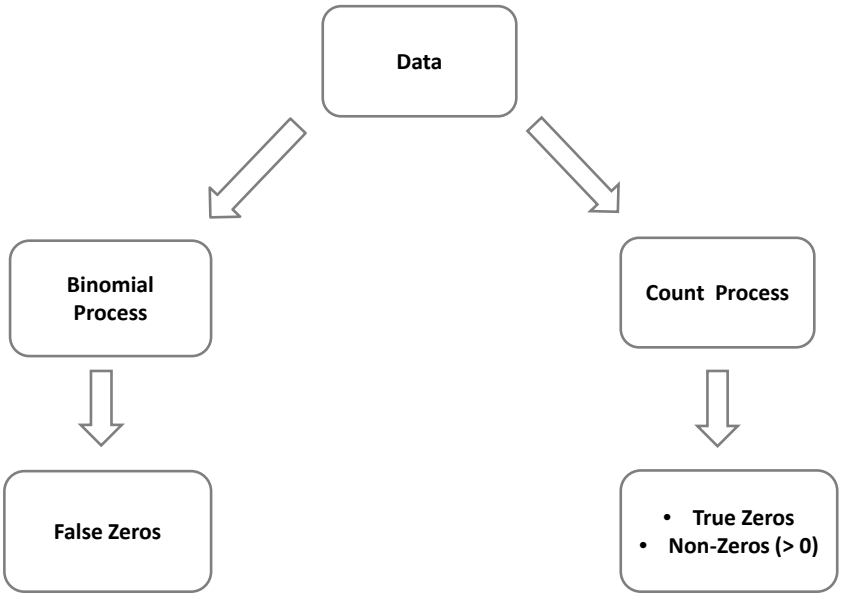


Figure 86: The main categories of the observed outcomes and the processes responsible for generating them.

We, however, note at this point that these categories are only imaginary and there is actually no way of distinguishing between the observed zero outcomes. In the following section, we base the formulation of the optimal model for our data on these two categories of zeros as well as the observed non-zero outcomes. We draw most of our concepts from Zuur et al. 2011.

2.5.1.3.1 Probability of Observing A Zero Outcome

It follows from Figure 86 that each time a negative urine slide is read, the zero outcome could either be a false or true zero. So, what then would be the probability of observing a zero outcome at all? According to Zuur et al. 2011, the probability of observing a zero outcome, $\texttt{Pr(Y}_i = 0)$, is given by equation 2. The second and third lines of equation 2 are the results of assuming that `Pr (False zeros)` and `Pr (Count process gives a zero)` follow a binomial and negative binomial probability distributions, respectively.

$$\Pr(Y_i = 0) = \begin{cases} \Pr\text{ (False zeros)} + \Pr\text{ (1 - Pr (False Zeros))} \times \Pr\text{ (Count process gives a zero)} \\ \pi_i + (1 - \pi_i) \times \Pr\text{ (Count process gives a zero)} \\ \pi_i + (1 - \pi_i) \times \left(\frac{k}{\mu_i + k}\right)^k \end{cases} \quad (2)$$

where `Pr(False zeros)` and `Pr(1 - Pr(False Zeros))` are the probabilities that the observed outcome is and is not a false zero, respectively; while `Pr(Count process gives a zero)` is the probability that we observe a true zero. The negative binomial distribution function for `Pr(Count process gives a zero)` in equation 2 was deprived as:

$$f(0; k, \mu_i) = \frac{\Gamma(0 + k)}{\Gamma(k) \times \Gamma(0 + 1)} \times \left(\frac{k}{\mu_i + k}\right)^k \times \left(1 - \frac{k}{\mu_i + k}\right)^0 = \left(\frac{k}{\mu_i + k}\right)^k \quad (3)$$

2.5.1.3.2 Probability of Observing A Non-Zero Outcome

On the other hand, the probability of observing a non-zero outcome, `Pr(`$Y_i > 0$`)`, is given by equation 4. As for equation 2 above, the second and third lines of equation 4 are obtained by assuming a binomial and negative binomial probability distributions for `Pr` (1 - Pr(False zero)) and `Pr (Count process)`, respectively.

$$\Pr(Y_i > 0) = \begin{cases} \Pr(1 - \Pr\text{ (False zero)}) \times \Pr\text{ (Count process)} \\ (1 - \pi_i) \times \Pr\text{ (Count process)} \\ (1 - \pi_i) \times \frac{\Gamma(y_i + k)}{\Gamma(k) \times \Gamma(y_i + 1)} \times \left(\frac{k}{\mu_i + k}\right)^k \times \left(1 - \frac{k}{\mu_i + k}\right)^{y_i} \end{cases} \quad (4)$$

where the gamma function, Γ, is given by $\Gamma(y+1) = y!$ whilst the second and third components of the negative binomial probability distribution function are power functions of the two unknown parameters that characterise the negative binomial distribution, the mean (μ) and dispersion parameter (k) (Zuur et al. 2013).

2.5.1.3.3 Modelling the Mean of the Count Process

By convention, the mean of the count and binomial processes are modelled separately. This section, therefore, focuses on the modelling of the count process. The mean of the count process, which in our case represents the mean burden of schistosome infections harboured per individual inhabitant, was modelled as a function of the measured covariates, $\mathbf{X}$ (equation 5).

$$\mu_i = e^{\alpha + \mathbf{X}_i \beta} \tag{5}$$

where α denotes the intercept, $\mathbf{X}_i$ denotes the vector of measured covariates and β is the vector of unknown regression coefficients.

2.5.1.3.4 Modelling the Mean of the Binomial Process

The binomial process was modelled as the probability of observing a false zero, π, versus any other observed outcomes (equation 6). From our discussion on the possible sources of the false zeros in this study, section 2.5.1.1, two of the measured covariates that may influence the probability of observing a false zero are older age and treatment history. This is because chronic morbidities such as fibrosis of the bladder, that increase schistosome egg retention rate in the infected human hosts; are the results of prolonged exposure, rather than the current burden of infection (Medley & Bundy 1996).

$$\pi_i = \frac{e^{\nu}}{1 + e^{\nu}} \tag{6}$$

Therefore, the middle-aged inhabitants, for instance, had a higher chance of excreting urine that did not contain any schistosome eggs despite being infected. Moreover, individual health-seeking behaviours could also lead to decreased schistosome burden, and consequently decreased egg excretion rate. We, therefore, chose to model the binary process as a function of older age and individual treatment history (equation 7).

$$\pi_i = \frac{e^{\nu} + \mathrm{X}_{1_i}\gamma_1 + \mathrm{X}_{2_i}\gamma_2}{1 + e^{\nu} + \mathrm{X}_{1_i}\gamma_1 + \mathrm{X}_{2_i}\gamma_2} \tag{7}$$

where ν and γ denote the intercept and the regression coefficient, respectively; whilst X_{1_i} and X_{2_i} denote the covariates for older age and individual treatment history, respectively.

2.5.1.3.5 Estimating the Unknown Regression Parameters

Our model has been formulated up to the unknown regression parameters $(\alpha, \nu, \beta, \gamma)$. The next stage of the analysis was, therefore, to estimate these unknown regression parameters using our data. As for other generalised linear models, the regression parameters of a zero-inflated negative binomial model are estimated by the method of maximum likelihood (Hilbe 2011). Using the data and the chosen model, the maximum likelihood method estimates values of the regression parameters that make the data most likely (Crawley 2014). The estimation process, therefore, involves the specification a joint likelihood criterion for the data which is then maximised as a function of the regression parameters (Zuur et al. 2013).

Conventionally, the parameter estimation process involves five main stages. The process begins with the formulation of the likelihood function based on the probability distribution function. The logarithm of the likelihood function is then taken in order to simplify the maximisation process. Therefore, the likelihood criterion is made additive from this point onwards. The estimation of values of the regression parameters that maximise the log-likelihood is attained by the frequentist approach which optimises the log-likelihood by obtaining the first order derivatives which are then set to zero. The parameter estimates and their corresponding standard errors are then attained by using a suitable optimisation routine to solve the resulting equations (Zuur et al. 2013, Hilbe 2011). To fit the model, we employed the use of the `zeroinfl` function, as implemented with the R `pscl` package, and the gradient was estimated by the BFGS algorithm.

2.5.1.4 Choosing the Optimal Model

Our formulation of a zero-inflated negative binomial model is based on the excess zero counts in our study outcome (Figure 84). However, given the fact that our study outcome is consistent with the distributional assumptions of the negative binomial probability function, we may have as well fitted a traditional negative binomial model. But

would such a model have provided an adequate fit to the data? Therefore, before proceeding with the model selection for the specified zero-inflated model, we conducted an assessment to determine if a traditional negative binomial model, rather than the zero-inflated model would have provided a more adequate fit to the data. The two models were, therefore, fitted as equations 8 and 9 using specific functions in the `R MASS` and `pscl` packages, respectively (Venables & Ripley 2002, Zeileis et al. 2008).

$$\begin{gathered} Y_i \sim NB(\mu_i, k) \\ E(Y_i) = \mu_i \qquad \text{and} \qquad var(Y_i) = \mu_i + \frac{\mu_i^2}{k}. \\ log(\mu_i) = \alpha + \beta \mathbf{X}_i \end{gathered} \tag{8}$$

$$\begin{gathered} Y_i \sim ZINB(\mu_i, \pi_i, k) \\ E(Y_i) = (1 - \pi_i) \times \mu_i \qquad \text{and} \qquad var(Y_i) = (1 - \pi_i) \times \mu_i \times (1 + \pi_i \times \mu_i + \frac{\mu_i^2}{k}) \\ log(\mu_i) = \alpha + \beta \mathbf{X}_i; \qquad logit(\pi_i) = \nu + \gamma_1 \mathrm{X}_{1_i} + \gamma_2 \mathrm{X}_{2_i} \end{gathered} \tag{9}$$

where μ_i, π_i, α, β and ν have the same meanings as before, $\mathbf{X}_i$ is the vector of measured covariates for the i^{th} participant whilst X_{1_i} and X_{2_i} denote the specific covariates whose effect were adjusted for in the binary component of the zero-inflated model. The expectation and variance for the observed outcome in the i^{th} participant, Y_i, are given by $E(Y_i)$ and $var(Y_i)$, respectively.

2.5.1.4.1 Assessing the Comparative Fit of the Two Models

The Vuong test, as implemented with the `pscl` package, was used in assessing the comparative worth of the two non-nested models (equations 8 and 9) (Zeileis et al. 2008, Hilbe 2014). The test, which is given by equation 10, is based on the null hypothesis that the two models are indistinguishable from each other. Therefore, the predicted probabilities of the two models are compared under the assumption that the test statistic is normally distributed, $N(0, 1)$. Hence, at the 95% confidence level, values of $V > 1.96$ would indicate a preference for the traditional negative binomial model whilst values of $V < -1.96$ would indicate a preference for the zero-inflated model (equation 11) (Hilbe

2011). The results of the test for our models provided overwhelming evidence in favour of the zero-inflated model over the traditional non-inflated model ($p = 0.007$, test statistic $= -2.46$).

$$V = \frac{\sqrt{n}\bar{u}}{SD(u_i)} \tag{10}$$

$$u_i = \ln\left(\frac{\sum_i \mathrm{P_{NB2}}(y_i|\mathrm{x}_i)}{\sum_i \mathrm{P_{ZINB}}(y_i|\mathrm{x}_i)}\right) \tag{11}$$

where $\mathrm{P_{NB2}}(y_i|\mathrm{x}_i)$ is the predicted probability of y given x by the traditional negative binomial model (equation 8); and $\mathrm{P_{ZINB}}(y_i|\mathrm{x}_i)$ is the predicted probability of y given x by the zero-inflated negative binomial model (equation 9); n denotes the number of observations whilst the mean and standard deviations of u are given by $\bar{u}$ and SD(u), respectively (Hilbe 2011).

We, however, note that the Vuong test is also known to be biased towards zero-inflated models since the same data is used in estimating parameters in both the count and binary components models (Hilbe 2014). Though a number of correction tests are often used in conjunction with the Vuong test to remedy this bias, none of those tests were available in `R` as of the time of preparing this report. Therefore, our results of the Vuong test may come with some amount of bias.

2.5.1.5 Interaction Terms

Table 54 outlines the interactions terms that were considered but whose effect could not be explored.

2.5.1.6 Model Selection

The effects of the individual terms in the maximal model was assessed with the likelihood ratio test. We, however, first needed to maintain a constant value for the dispersion parameter, k, which the `pscl` package refers to as θ; in the full and nested models for the computation of the likelihood ratio statistic. Hence, in doing so we adapted a function by Zuur et al. 2012 which is based on the `pscl`'s *zeroinfl* function with the

Table 54: Interaction terms that were considered but whose effects could not be explored.

Interaction Term	Reason For Considering The Effect	Reason(s) Why It Was Impossible To Explore The Effect
Age : Treatment	To explore the re-infection rates following treatment in the younger and older age groups.	1. The community was a pre-intervention site 2. The 22 participants who had sought treatment in the year preceding our survey were mostly under the age of 16. Hence, there was no basis for the comparison.
Age: Frequency of Contact	To explore the varying exposure patterns with age	The available information on the frequency of contact was limited by its lack of the exact number of contact per person on daily basis.
Sex: Frequency of Contact	To explore the role of gendered tasks on varying exposure patterns.	No available information on the exact number of water contact per person on daily basis.

following modifications: the number of parameters to be estimated by the optimisation routine is reduced by 1; and the value of θ is pre-set to its value in the full model.

The computation of the likelihood ratio statistic, equation 12, was therefore effected by first dropping the model terms sequentially from the maximal model. The differences between the log-likelihoods of the nested and full models were then compared with critical values of the χ^2 - distribution. Using the results of the likelihood ratio test, a backwards selection was performed by the sequential omission of the least significant terms from the full model. The resulting model, therefore, became our provisional model for the aggregated schistosome burden.

$$-2\left[\mathrm{LL}_{\mathrm{reduced}} - \mathrm{LL}_{\mathrm{full}}\right] \tag{12}$$

where LL is the log-likelihood. The computed difference between the two models is Chi-squared distributed with degrees of freedom given by the difference in number of predictors (Hilbe 2009).

2.5.1.7 Goodness-Of-Fit Check

2.5.1.7.1 Assessment of Over-dispersion

The Pearson dispersion statistic, as given by equation 13, was computed to assess the provisional model for over-dispersion. In line with suggestions by Hilbe 2011, we set our cut-off limit for over-dispersion at 1.25.

$$\delta = \frac{\chi^2}{df_1} \tag{13}$$

where

$$\chi^2 = \phi \times \sum_{i=1}^{\mathrm{N}} \frac{(\mathrm{Y}_i - \mathrm{E}(\mathrm{Y}_i))^2}{var(\mathrm{Y}_i)}$$

$$df_1 = \mathrm{N} - p$$

where χ^2 denotes the sum of squared Pearson residuals; df_1 is the residual degrees of freedom; N is the number of observations and p is the number of parameters in the model, including the intercept (Zuur et al. 2013).

2.5.2 Model Validation

2.5.2.1 Assessing the Homoscedasticity Assumption

As part of the model validation, we conducted an assessment to check for any obvious non-linear patterns in the standardised residuals, equation 14, of the provisional model. The improper modelling of a covariate effect or the omission of relevant covariates from the model are the common sources of residual patterns (Zuur et al. 2012). Therefore, any such patterns in the residuals signify a violation of the homogeneity assumption, $\epsilon_i \sim N(0, \sigma^2)$; and hence the need to further improve the model. We, therefore, conducted a graphical assessment of the count model's Pearson residuals against their corresponding fitted values as well as the Pearson residuals against each of the covariates in the provisional model. The former plot was additionally used in identifying observations that potentially had influential effects in the model and which may have influenced the value of the Pearson dispersion statistic (equation 13).

$$\epsilon_i = \frac{Y_i - (1 - \pi_i) \times \mu_i}{\sqrt{var(Y_i)}} \tag{14}$$

where

$$E(Y_i) = (1 - \pi_i) \times \mu_i$$

$$var(Y_i) = (1 - \pi_i) \times \mu_i \times (1 + \pi_i \times \mu_i + \frac{\mu_i^2}{k})$$

where Y_i denotes the observed value of the study outcome for the i^{th} participant; $E(Y_i)$ and $var(Y_i)$ are respectively the expectation and variance of Y_i.

2.5.2.2 Assessing the Assumption of Independence

Going back to our primary study objective, we set out to model the risk of aggregated schistosome burden as a function of the measured covariates as well as a number of potentially relevant unmeasured factors that may have influenced the levels of infectivity of the water contact sites (Table 52 page 194). However, while our provisional model is based on the measured covariates, the effect of the unmeasured factors still remains unaccounted for.

Therefore, as part of the model's assessment of the assumption of independence of observations, which is violated by the omission of important covariates, we will be focussing on the unmeasured factors in the rest of this section. Though the unmeasured factors under consideration were closely linked to the littoral zone and hence the ecology of the intermediate snail hosts, the mechanism by which they took effect in the human host population involved the exposure of different subsets of the inhabitants to specific water contact sites along the shoreline (Figure 87).

The choice of specific water contact sites by members of any particular household in our study community, however, depended on the proximity of that household to the water contact site as well as the purposes for which the sites were reserved. Moreover, households in the community mainly comprised of kinship loci and members of the same clans tend to live in close proximity to each other. Consequently, members of any of these familial aggregations in the community mainly carried out most of their exposure activities, especially the ones that involved domestic chores, at the same water contact sites.

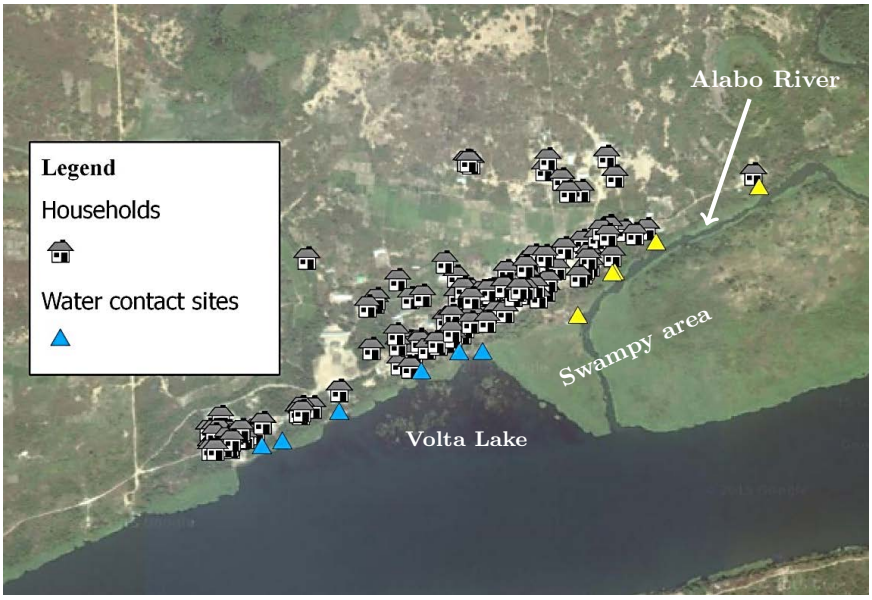


Figure 87: Distribution of water contact sites in our study community. The blue-coded sites are situated along the main shoreline of the Volta Lake whilst the yellow-coded sites are along the banks of the Alabo River, a tributary of the Volta Lake.

Therefore, judging by the fact that residual spatial effects manifest as spatial autocorrelation (Arlinghaus 1996), we could logically assess the effect of the unmeasured factors under consideration as latent patterns in the point-referenced residuals. Moreover, given the heterogeneities in exposure patterns across the different households in the community, we would expect the varying likelihood in the risk of our study outcome, aggregated schistosome burden, to vary across households depending on the levels of infectivity of the water contact sites that were frequented by the members of the various households.

2.5.2.2.1 Water Contact Sites

We have so far been assuming that exposure patterns were diffused across all the water contact sites in the our study community. Therefore, we expect the unmeasured factors under consideration, which took effect through frequent exposure, to manifest in the degree of schistosome burden harboured by the different subsets of inhabitants who frequented the various water contact sites. However, in order for this assumption to be plausible, the transmission of infections should ideally occur in all of the community's water contact sites.

As evident from Figure 87, however, the community had two categories of water contact sites: those along the main shoreline of the lake; and those along the banks of the Alabo River. While the transmission of schistosome infections in the lake is well-established (Dalton & Pole 1978, Scott et al. 1982, Chu et al. 1981), there were no records of transmissions occurring in the river.

However, the transmission of schistosome infections in any freshwater body is characterised by the presence of the snail host species as well as a relatively stagnant velocity of flow of the water body and the presence of the emergent plant, *Ceratophyllum sp* (Cowper 1971, Klumpp & Chu 1980). We, therefore, used the two latter factors, Figure 88, as basis for ascertaining that the Alabo River was a likely source of transmissions in the community. Hence, "lake" and "river" are used interchangeably in the remainder of this text.

(a)

(b)

(c)

Figure 88: The flow rate of the Alabo River during the period of stagnation (a) and the raining season (b). *Ceratophyllum sp* (arrowed), in the river (c).

2.5.2.2.2 Assumptions Regarding the Effects of the Unmeasured Factors

The residuals are regarded as the variation in the response that is associated with the effect of unobserved factors as well as the random noise in the data (Arlinghaus 1996). Hence, assessing the effect of the unmeasured factors as latent spatial patterns in the point-referenced residuals would mean the following assumptions are being made: 1) that the unmeasured factors are spatially varying; 2) that their collective effect plays a relevant role in explaining the community-level variation in the risk of aggregated schistosome burden; and 3) that the residuals of our provisional model would be spatially correlated due to the omission of these unmeasured factors from the model (Waller & Gotway 2004).

Scientifically, it would also mean that the subset of inhabitants who harboured aggregated schistosome burden would be expected to have experienced their exposures at the water contact sites that were most infective and vice versa. Moreover, since members of the same familial aggregations lived in close proximity to each other and frequented common water contact sites, we would ideally expect the varying likelihood in risk to cluster by these familial aggregations across the community.

Figure 89: A group of men transplanting rice. Working on rice plantations is considered a high risk occupation as far as the transmission of schistosome infections is concerned (Cowper 1971). Image source: MoFA 2015

We would also be assuming that all *S. haematobium* transmissions were autochthonous. Therefore, exposures that occurred outside the community's shoreline were assumed to have had negligible effect on the community-level variation in the risk of aggregated schistosome burden. The most notable of these "external" exposures involved working on rice plantations in a community across the lake (Figure 89).

The next stage of the analysis, therefore, involves the sorting out of the model into covariate information and residuals. But first, we revisit some concepts in probability and classical geostatistics that are of relevance.

2.5.2.3 Classical Geostatistical Concepts

Since the community-level variation in the risk of our study outcome may be influenced by a myriad of factors, most of which may still be largely unknown, we could adopt a stochastic view at this point and regard the outcome of these factors as random. However, adopting a stochastic view would also imply that each sampled location within our study community would have a set of possible outcomes, rather than just the one observed value. Therefore, our observed outcome could be regarded as one drawn at random from a set of possible outcomes (Webster & Oliver 2001).

Hence, following classical geostatistical convention, we began by assuming that the observed outcome at each sampled location, $\mathbf{s}$, within the study community were samples of a single realisation, $z(s_i)$, of an underlying random and spatially continuous process, $\mathbf{Z}$. So, in effect each sampled location could be regarded as having a population value, denoted by $\mathbf{Z}(s_i)$, and the actual observed realisation of $\mathbf{Z}(s_i)$ given by $z(s_i)$. Therefore, in order to make statistical inferences from the observed realisation, $\mathbf{Z}$ is conventionally specified by its mean, μ, and covariance or variogram (equation 15) (Gelfand et al. 2010, Waller & Gotway 2004).

$$E\left[\mathbf{Z}(\mathbf{s})\right] = \mu \tag{15}$$

$$Cov\Big(\mathbf{Z}(\mathbf{s}_j), \mathbf{Z}(\mathbf{s}_k)\Big) = C(\mathbf{s}_j - \mathbf{s}_k)$$

where $C(.)$, the covariance function, measures the spatial autocorrelation between sampled locations $\mathbf{s}_j$ and $\mathbf{s}_k$. Under the assumption of second-order stationarity for the random

function, $\mathbf{Z}$, μ is independent of location and the covariance only depends on the separation distance between $\mathbf{s}_j$ and $\mathbf{s}_k$(Waller & Gotway 2004).

So, building on the assumptions in section 2.5.2.2.2, we expect the residual spatial variation due to the unmeasured covariates to manifest as spatial autocorrelation (Waller & Gotway 2004). However, since the range of unobserved factors in the causal pathway of our study outcome may not be limited to the unmeasured factors under consideration; any observed spatial trend in the residuals may not be entirely attributable to the effect of the unmeasured factors under consideration.

In the next stage of the analysis, the variogram is employed in assessing the spatial correlation between point-referenced residuals, $\mathrm{z}(s_i)$, for neighbouring households in this study; and to describe any such spatial dependence with the appropriate correlation structure.

2.5.2.4 Structural Analysis

Under the assumption of intrinsic stationarity for the Pearson residuals i.e. $\epsilon_i \sim N(0, \sigma^2)$, an omni-directional empirical variogram, equation 16, was computed to quantify the second-order dependence in the residuals (Diggle & Ribeiro 2007). Full details of these computations have already been presented in Part I of this thesis.

$$\gamma(\mathbf{h}) = \frac{1}{2p(\mathbf{h})} \sum_{\alpha=1}^{p(\mathbf{h})} \{z_i(\mathbf{s}_j) - z_j(\mathbf{s}_k)\}^2 \tag{16}$$

where $p(\mathbf{h})$ denotes the number of pairs of Pearson residuals of the provisional model that are separated by $\mathbf{h}$; the spatial lag, $\mathbf{h}$, is the distance separating any given set of pairs of residuals; whilst z_i and z_j are the residuals for sampled locations $\mathbf{s}_i$ and $\mathbf{s}_j$, respectively.

2.5.2.4.1 Assessing Directional Dependence

The omni-directional variogram is based on the assumption that the spatial correlation only depends on the Euclidean distance between sampled locations. However, we are hypothesising that the risk of aggregated *S. haematobium* burden may vary across our study community depending on which water contact sites were frequented by members of the different familial aggregations. Therefore, if the different water contact sites did

indeed exhibit heterogeneities in their levels of infectivity, then there was the possibility that the spatial correlation would be stronger in the direction of those sites that contributed most to the transmission potential of infection in the community, assuming autochthonous transmission. Hence, we employed the use of anisotropic variograms in investigating if the spatial dependence may have varied with the relative orientation of the sampled locations. These anisotropic variograms were then examined for differences in the slope of their curves that would suggest a violation of the isotropy assumption and hence, the need to correct for the anisotropy (Diggle & Ribeiro 2007, Waller & Gotway 2004, Jr & Diggle 2016). The methods for correcting for anisotropy have already been discussed in section 3.2.2.2.3 in Part II of this thesis.

An envelope of random permutations, together with a formal test for trend, was used in assessing if any observed trend in the isotropic empirical variogram may have occurred by chance (Diggle & Ribeiro 2007, Eagle & Diggle 2012).

2.5.2.5 Interpretation of the Formal Test for Trend

2.5.2.5.1 Lack of Significant Spatial Trend in the Residuals

The formal test for trend (Eagle & Diggle 2012), which is based on a null hypothesis of the absence of spatial autocorrelation in the residuals, was used as basis for assessing the relevance of the effect of the unmeasured factors under consideration in this study. Hence, the absence of significant spatial autocorrelation was interpreted as signifying that the effect of the unmeasured factors could be disregarded without influencing the validity of our classical zero-inflated negative binomial model (Arlinghaus 1996). However, it is worth noting that the absence of significant spatial autocorrelation could also mean that the effect of the unmeasured factors was very weak. This is because the decomposition of the data into large and small scale variations tend to depend on the strength of the covariate effects as well as the amount of variability in the data (Waller & Gotway 2004).

2.5.2.5.2 Presence of Significant Spatial Trend in the Residuals

On the other hand, significant spatial autocorrelation in the residuals is indicative of residual spatial variation, and hence an over-estimation of the provisional model's pre-

cision due to the lack of independent information on the autocorrelated observations (Guelat 2013). This would also mean that the unmeasured covariates may have indeed played a relevant role in explaining any observed community-level variation in the risk of aggregated schistosome burden.

Therefore, under such circumstances, adapting the provisional model to allow for the spatially correlated errors would be expected to "mop up" the residual spatial variation due to the unmeasured covariates under consideration as well as any other unobserved factors that may have played relevant roles in the causal pathway of our study outcome (Waller & Gotway 2004).

For instance, suppose we fit a model with the measured covariates, $\mathbf{X}(\mathrm{s}_i)$, and the spatially correlated residuals, $W(s_i)$ (equation 17). If in reality, there is a single unmeasured covariate, $x^*(s_i)$, with a regression parameter, β^*, and uncorrelated residuals, then the model would be given by equation 18. Therefore, if these two models were compared, the spatially correlated process, $W(s_i)$, would act as a proxy for $x^*(s_i)\beta^*$. The exact extent of this comparability would, however, depend on the strength of the effect of the unmeasured covariate (Waller & Gotway 2004). A detailed account of the geostatistical methodology for adjusting for the effect of unobserved covariates has been discussed in sections 2.8.3.3 - 2.8.3.5 of Part II of this thesis.

$$\begin{aligned} log(\mu_i) &= \alpha + \beta\mathbf{X}(s_i) + W(s_i) + \epsilon_i \\ logit(\pi_i) &= \nu + \gamma_1 X_1(s_i) + \gamma_2 X_2(s_i) + \xi_i \end{aligned} \tag{17}$$

$$\begin{aligned} log(\mu_i) &= \alpha + \beta\mathbf{X}(s_i) + x^*(s_i)\beta^* + \epsilon_i \\ logit(\pi_i) &= \nu + \gamma_1 X_1(s_i) + \gamma_2 X_2(s_i) + \epsilon_i \end{aligned} \tag{18}$$

3 Results

3.1 Community-Level Variation in the Points of Exposure

Exposure activities in the community were distributed across 11 water contact sites along the shoreline (Figure 90). However, judging by our questionnaire responses on the point of regular exposure, Tables 55 – 56, the Akpeakpe and Tutukope water contact sites seemed to have been the focus of maximum exposure activity. Therefore, assuming transmissions were autochthonous, these two sites may have contributed greatly to the overall transmission potential of infections in the community.

However, this study is also based on the hypothesis that the different water contact sites, which were mainly reserved for different purposes, may have varied in their levels of infectivity due to the cumulative effect of the site-specific anthropogenic activities. Therefore, building on our earlier concepts, a closer examination of some of the site-specific exposure activities in the community would suffice to illustrate how the natural ecology of the intermediate snail hosts may have been altered positively or otherwise.

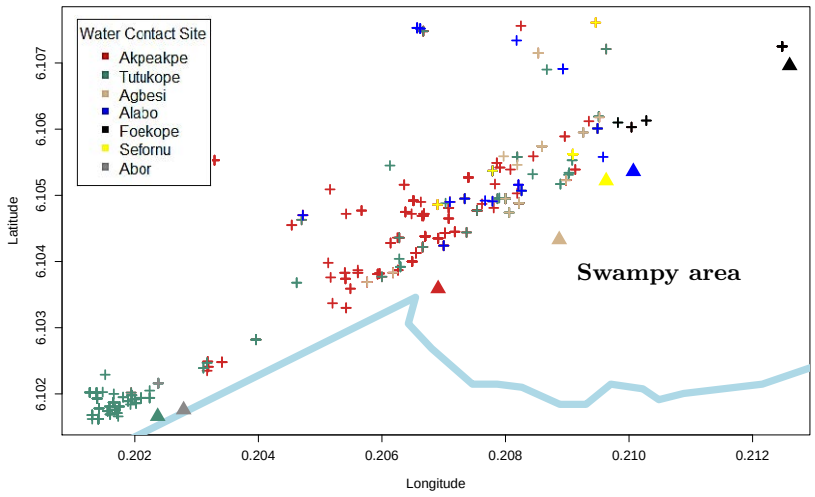


Figure 90: Distribution of households (+) colour-coded by the water contact sites frequented (Δ) by members. The last four water contact sites that appear beyond the main shoreline were situated along the banks of the Alabo River, a tributary of the Volta Lake. These sites were, however, accessible by canoes from the main shoreline through the swampy area.

Table 55: Intensity of infection among the female inhabitants and the water contact sites they mostly frequented.

Infection Status	Age-Group	Total	Water Contact Sites							
			Abor	Agbesi	Akpeakpe	Alabo	Ameworgbe	Foekope	Sefornu	Tutukope
Heavily Infected	5-12 years	24	1 (4.17)	0 (0.00)	17 (70.83)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	6 (25.00)
	13-38 years	8	0 (0.00)	2 (25.00)	4 (50.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (25.00)
	39-51 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	52-60 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Lightly Infected	5-12 years	42	0 (0.00)	2 (4.76)	30 (71.43)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	10 (23.81)
	13-38 years	28	0 (0.00)	0 (0.00)	16 (57.14)	2 (7.14)	0 (0.00)	1 (3.57)	0 (0.00)	9 (32.14)
	39-51 years	3	0 (0.00)	0 (0.00)	1 (33.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	2 (66.67)
	52-60 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Non-Infected	5-12 years	63	0 (0.00)	2 (3.17)	27 (42.86)	5 (7.94)	0 (0.00)	1 (1.59)	6 (9.52)	22 (34.92)
	13-38 years	105	1 (0.95)	3 (2.86)	51 (48.57)	2 (1.90)	0 (0.00)	5 (4.76)	0 (0.00)	43 (40.95)
	39-51 years	44	0 (0.00)	1 (2.27)	23 (52.27)	2 (4.54)	0 (0.00)	2 (4.54)	0 (0.00)	16 (36.36)
	52-60 years	14	0 (0.00)	0 (0.00)	10 (71.43)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	4 (28.57)
	52-60 years	5	0 (0.00)	2 (40.00)	1 (20.00)	2 (40.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Total		331	2 (0.60)	10 (3.02)	179 (54.08)	11 (3.32)	0 (0.00)	9 (2.72)	6 (1.81)	114 (34.44)

Table 56: Intensity of infection among the male inhabitants and the water contact sites they mostly frequented.

Degree of Infection	Age-Group	Total	Water Contact Sites							
			Abor	Agbesi	Akpeakpe	Alabo	Ameworgbe	Foekope	Sefornu	Tutukope
Heavily Infected	5-12 years	22	0 (0.00)	1 (4.55)	20 (90.91)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (4.55)
	13-38 years	17	0 (0.00)	0 (0.00)	11 (64.71)	0 (0.00)	0 (0.00)	0 (0.00)	1 (5.88)	5 (29.41)
	39-51 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
	52-60 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Lightly Infected	5-12 years	30	0 (0.00)	0 (0.00)	22 (73.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	8 (26.67)
	13-38 years	17	0 (0.00)	1 (5.88)	8 (47.06)	1 (5.88)	0 (0.00)	0 (0.00)	1 (5.88)	6 (35.29)
	39-51 years	0	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0(0.00)	0 (0.00)
	52-60 years	1	0 (0.00)	0 (0.00)	1 (100)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Non-Infected	5-12 years	47	0 (0.00)	4 (8.51)	20 (42.55)	4 (8.51)	0 (0.00)	0 (0.00)	1 (2.13)	18 (38.30)
	13-38 years	40	0 (0.00)	3 (7.50)	13 (32.50)	2 (5.00)	0 (0.00)	1 (2.50)	0 (0.00)	21 (52.50)
	39-51 years	12	0 (0.00)	1 (8.33)	6 (50.00)	0 (0.00)	0 (0.00)	1 (8.33)	0 (0.00)	4 (33.33)
	52-60 years	5	0 (0.00)	2 (40.00)	1 (20.00)	2 (40.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Total		191	0 (0.00)	12 (6.28)	102 (53.40)	9 (4.71)	0 (0.00)	2 (1.04)	3 (1.57)	63 (32.98)

3.2 Sources and Effects of Ecological Changes

The ecology and bionomics of the snail host species within their freshwater habitats are known to be influenced by a series of factors which can be categorised as physical, chemical and biological. Taken together, these categories tend to cover a wide range of factors that ultimately influence the breeding, availability of food sources and general survival of these snails (Cowper 1971). For instance, physical conditions that influence the velocity of flow of the aquatic habitats of these snails may adversely affect their survival by disturbing their breeding sites, washing away their food sources or dislodging them.

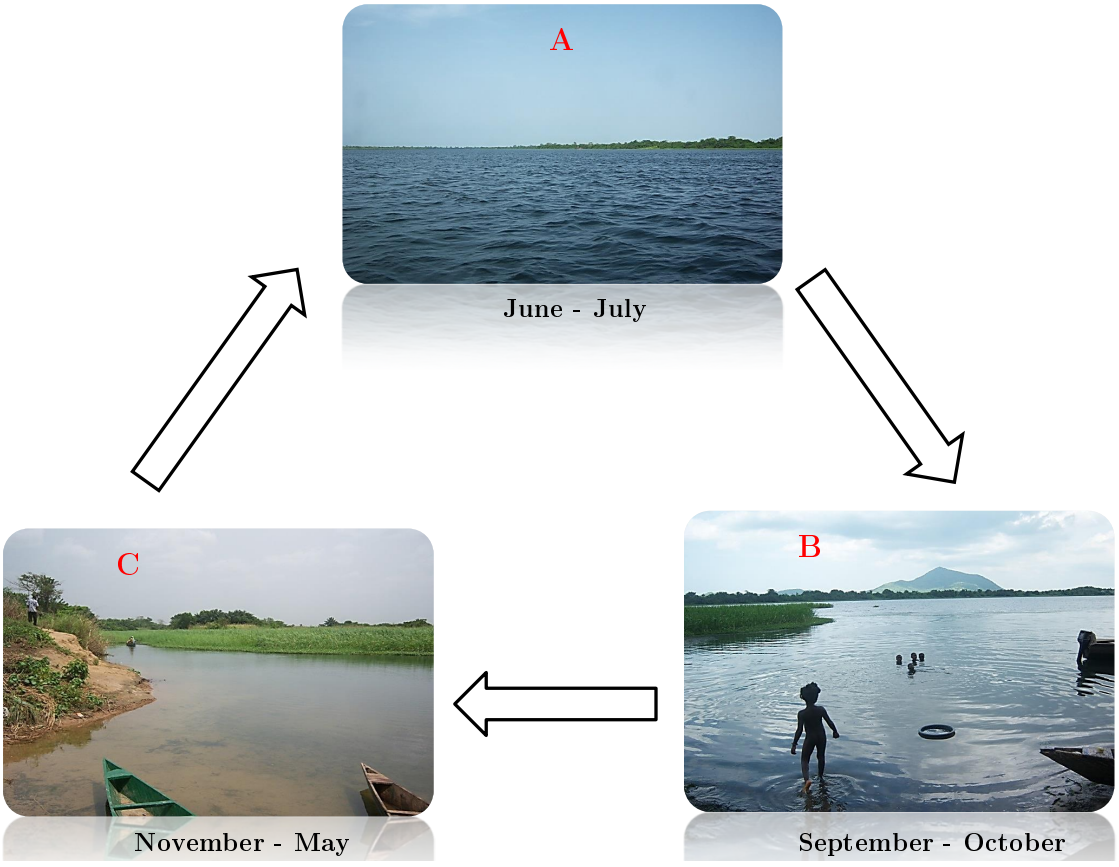


Figure 91: The annual cycle of the lake with emphasis on the velocity of flow. A: The high flow rate during the major rainy season in June-July when the spillways of the Akosombo Dam were opened six miles upstream; B: alternating periods of increased flow rate and stagnation during the minor rainy season in September-October; and C: the period of absolute stagnation between November-May.

However, while natural conditions such as floods or fluctuation in temperature may be responsible for most of these ecological changes, the resultant effect of certain anthropogenic activities in the littoral zone may also alter the ecology of these snails in various ways. The rest of this section will, therefore, focus on some of the specific purposes for which the different water contact sites in our study community were reserved and the potential effect of the site-specific exposure activities on the ecology of *Bulinus truncatus rohlfsi*, the intermediate snail hosts for *S. haematobium*.

We, however, examine the effect of these anthropogenic activities in the context of the annual cycle of the lake (Figure 91). Therefore, during the periods of stagnation, each water contact site was regarded as a discrete entity whose level of infectivity, based on the population of snail hosts it harboured, was a function of the direct consequence of the site-specific exposure activity on the ecology of these snails. Moreover, in order to establish a threshold point beyond which the effect of the site-specific anthropogenic activities became constant, we further assumed that the cumulative effect of the exposure activities was likely to have reached its peak around the middle of the stagnation period of the lake (Figure 91).

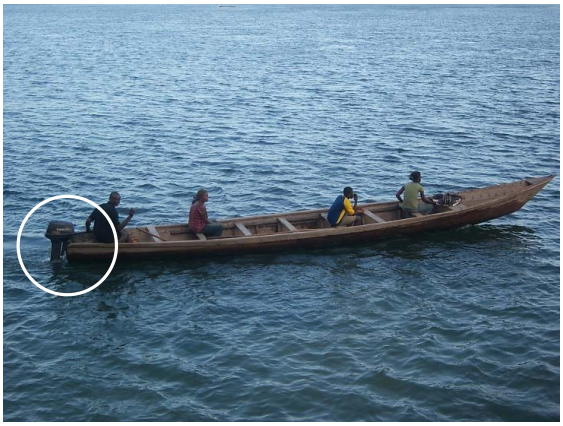
The annual cycle of increased flow rate of the lake in June-July was also assumed to dilute the effect of all exposure activities and cause fluctuations in the population density of the snail host species. Therefore, the water contact sites, which are rendered undefined by the high flow rate, could be regarded as unpolluted entities with a new generation of snail hosts at the end of every major rainy season.

3.2.1 Exposure Activities

3.2.1.1 Economic Purposes of Water Contact Sites

The normal lentic conditions of the littoral zone, that serve as habitats for the snail host species, may be altered by the effect of physical conditions such as the ripple action of waves, increased flow rate, floods or artificially-induced turbulence (Cowper 1971). However, since the emphasis in this case is on the period of stagnation of the lake, we would concentrate mainly on the sources and consequences of artificially-induced turbulence.

(a)



(b)

(c)

Figure 92: Manually-operated canoes (Figure 92(a)) and upgraded canoes fitted with outboard motors (Figure 92(b)). The Akpeakpe and Tutukope water contact sites also served as canoe berthing points (Figure 92(c)).

The effect of activities involving boats, motor-powered canoes and constant wading by humans are some of common sources of artificially-induced turbulence in the littoral zone (Cowper 1971). During the time of our pilot survey, almost all the canoes that operated along the shoreline were manually-powered (Figure 92(a)). However, by the time the main project took off a year later, most of the canoes that served commercial purposes had been upgraded with outboard motors (Figure 92(b)). Therefore, the two sites, Akpeakpe and Tutukope, that served as the community's canoe stations were constantly subjected to the activities of these motor-powered canoes; with potentially deleterious effects on the ecology of snail host species that inhabited these sites.

First of all, this induced turbulence may render the canoes' stations unsuitable for the snail hosts through the disruptions of their breeding sites or washing away of their food sources (Cowper 1971). Moreover, the fuel tanks of the outboard motors were usually filled by pouring gasoline directly from gallons while the canoes were berthed at the water contact sites; a technique which was prone to spillage. Therefore, oil slicks from these gasoline spillage were common on the surface of the water at the two canoe stations. These oil slicks, however, have the effect of interfering with the air supply of the snail hosts which respire by utilising atmospheric air directly from the surface of the water (Cowper 1971).

3.2.1.2 Domestic Purposes

The effect of chemical pollution by soap is known to render the littoral zone uncongenial for both the snail hosts and the larval stages of the schistosome worm (Okwuosa & Osuala 1993, Cowper 1971). While the washing of clothes and bathing may potentially result in chemical pollution, the setting-specific manner in which these activities are carried out may eventually determine the level of pollution associated with them. For instance, the inhabitants of our study community mainly bathed directly inside the lake. Therefore, some amount of chemical pollution, attributable to bathing, could be assumed for all the water contact sites along the shoreline.

(a)

(b)

(c)

Figure 93: Site-specific variations in the washing of clothes. (Figure 93(a)): the Agbesi water contact site (Figure 93(b)): the Abor and Figure 93(c): Tutukope.

However, the manner in which the washing of clothes was carried seemed to have varied from site to site. For instance, women who lived near the Tutukope water contact site seemed to prefer washing on their compounds while those who used the Alabo water contact site mainly stood on the shores while they washed. Yet still, women were observed in other sites washing directly inside the littoral zone (Figure 93). Therefore, the chemical pollution associated with washing may only prove effective in the latter case.

3.2.1.3 Recreational Purposes

The lake served as a source of recreation for children in the community. Therefore, swimming was a common group activity among children in the late afternoons (Figure 94). Empirical evidence, however, points to the fact that contact with cold water tends to stimulate the act of urination in humans (Farooq & Mallah 1966); and since children hardly control this urge, swimming could be associated with the pollution of the water contact sites by human excreta. It has also been consistently shown that pollution by human excreta is favourable for the survival of the snail host species, though the exact mode of utilisation is largely unknown (Cowper 1971).

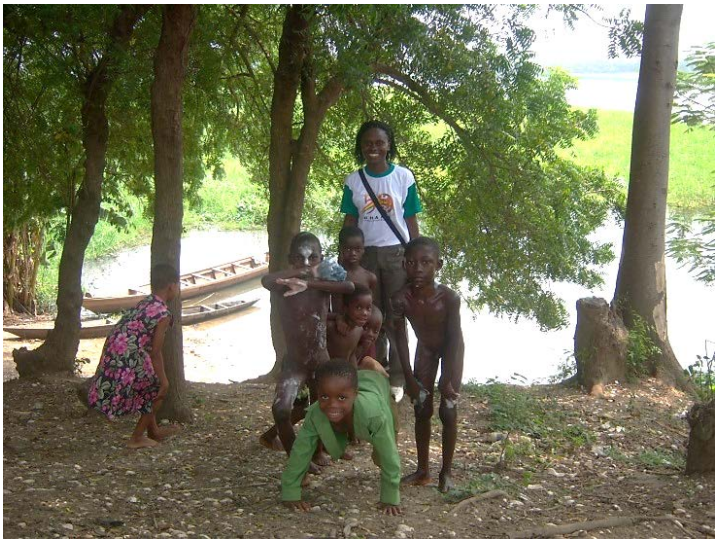


Figure 94: A group of children taking a break from their swimming at the Alabo water contact site. Swimming was a common recreational activity among children in the community.

3.2.2 Exploratory Analysis

3.2.2.1 Assessing Non-Linearity in Age

Biologically, the burden of *S. haematobium* infection tends to exhibit different patterns of variation across the various age groups in endemic areas (Jordan & Webbe 1982), thereby resulting in non-linear trends across time. An initial exploration of our data, Figure 95, suggested that this non-linear effect in the burden of infection was indeed obvious, especially among participants aged ≤ 25 years. Therefore, fitting age as a linear term in our model was inadvisable, as that would have resulted in misspecification problems.

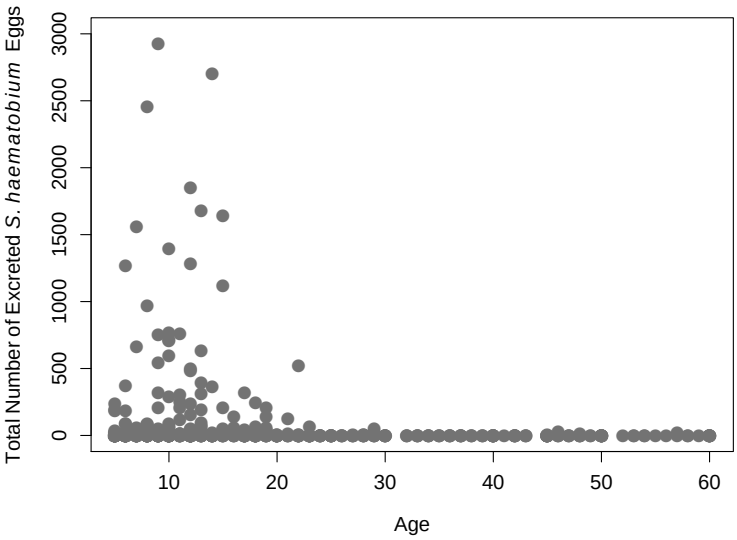


Figure 95: Scatterplot of *S. haematobium* egg count, in 10ml of urines, versus age of participants. The graph mainly suggests a non-linear effect in the burden of infection in participants aged below 25 years.

In order to ease the interpretation of the observed non-linear trend in Figure 95, we employ the use of non-parametric smoothers by Wood 2006 in the next section to estimate a curve that reasonably captures the trend across the age range for this study. We will then determine how best to incorporate the effect of age into our formal model for aggregated *S. haematobium* burden.

3.2.2.1.1 Estimated Non-Parametric Smoother for Age

In line with our earlier observations in section 3.2.2.1, the non-parametric smoother estimated a non-linear effect in age (Figure 96). Conventionally, the `gam` function by Wood 2006 estimates the smoother using thin plate regression spline; and by default, the

optimum amount of smoothing is also estimated automatically by cross-validation (Zuur et al. 2012). In our case, the cross-validation process produced 6.902 effective degrees of freedom, as opposed to the 1 degree of freedom that would have been produced for a linear effect (Table 57). In all, the smoother, which had a significant effect at the 5% level ($p<0.001$), accounted for 25% of the total variation in aggregated *S. haematobium* burden (Table 57).

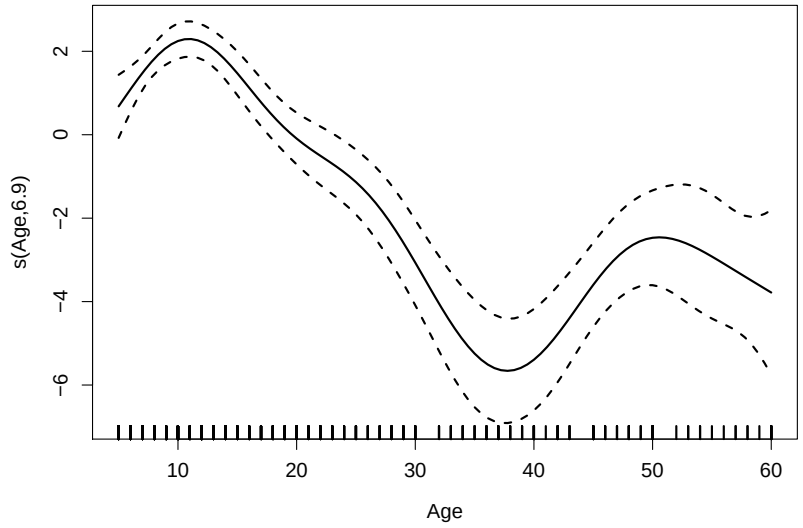


Figure 96: Estimated non-parametric smoother for age (solid curve) and the corresponding 95% confidence limits (black dashed lines). The smoother is plotted on the scale of the linear predictor and the value, 6.9, in the y-axis label represents the optimum amount of smoothing as determined by the cross-validation process.

3.2.2.1.2 Modelling the Effect of Age

With the evidence we have so far established in favour of non-linearity, the next logical step in the analysis was to incorporate this non-linear effect into our formal model. However, incorporating the smooth function directly into the model, equation 9, was not an option since the `pscl` package which was used in executing the model has not got the necessary features to directly cope with non-parametric smoothers (Zuur et al. 2012).

Therefore, instead of switching to another `R` package that copes with smoothers, such as `gamlss` by Rigby & Stasinopoulos 2005, to enable the fitting a semi-parametric model; we settled for the option of programming our own smoothing spline that could be executed with the `pscl` package.

Table 57: Model summary for the smooth function for age in the generalised additive model (Equation 1). By default, the *gam* function in the *mgcv* package used the Unbiased Risk Estimator (UBRE) score in the cross-validation process to determine the optimum amount of smoothing. The amount of smoothing is expressed as effective degrees of freedom (EDF) (Zuur et al. 2011).

Approximate Significance of Smooth Term					Fit Details				
Smooth Term	**EDF**	**Ref. df**	**Chi-Square**	**p-value**	**R-Square**	**UBRE score**	**Deviance explained**	**Scale est.**	**N**
s (Age)	6.902	7.918	247.5	0.000	0.0147	-0.11072	24.50%	1	522

3.2.2.1.3 Manual Programming of the Smoothing Spline

In line with our earlier discussion in section 3.1.2.1.3 in Part I of this thesis, the smoother, $f(\mathrm{X}_i)$, is defined by equation 19. The smoother was, therefore, regarded as a stick that could broken into p number of segments. By choosing our break-points to reflect the changes in the burden of infection, as estimated by our smoother at ages 12, 38 and 50 (Figure 96), we could determine the value of p. Therefore, using the broken stick model approach by White et al. 2014, the various segments of $f(\mathrm{X}_i)$ were defined by equation 20 - 21 and incorporated into our formal model.

$$f(\mathrm{X}_i) = \sum_{j=1}^{p} \beta_j \times b_j(\mathrm{X}_i) \tag{19}$$

where p is the total number of segments that make up $f(\mathrm{X}_i)$ and β_j is the regression coefficient obtained by fitting the j^{th} segment in the model.

$$\textbf{age1} = \begin{cases} \text{age}, & \text{if} \quad \text{age} \leq 12 \\ 12, & \text{if} \quad \text{age} > 12 \end{cases} \qquad \textbf{age2} = \begin{cases} 0, & \text{if} \quad \text{age} \leq 12 \\ \text{age} - 12, & \text{if} \quad 12 \leq \text{age} \leq 38 \\ 26, & \text{if} \quad \text{age} > 38 \end{cases} \tag{20}$$

$$\textbf{age3} = \begin{cases} 0, & \text{if} \quad \text{age} \leq 38 \\ \text{age}, & \text{if} \quad \text{age} > 38 \end{cases} \qquad \textbf{age4} = \begin{cases} 0, & \text{if} \quad \text{age} \leq 38 \\ \text{age} - 38, & \text{if} \quad 38 \leq \text{age} \leq 50 \\ 12, & \text{if} \quad \text{age} > 50 \end{cases} \tag{21}$$

3.2.3 Confirmatory Analysis

3.2.3.1 Model for Aggregated *S. haematobium* Burden

3.2.3.1.1 The Maximal Model

Tables 58 and 59 present the relative effect of the terms in the maximal models. The mean burden of schistosomes harboured by the individual inhabitants was modelled as a function of all the measured covariates in the count model (Table 58). On the other hand, the probability that an observed zero schistosome egg count was a false zero was modelled as a function of age3, age4 and individual treatment history with regard to seeking treatment for the symptoms of acute schistosomiasis with over-the counter medication (Table 59). Our reasons for adjusting the binomial model for these particular covariates have already been discussed in section 2.5.1.3.4.

Table 58: Maximal model for the mean burden of *S. haematobium* infections harboured per individual participant in our study community.

	Count Model			Likelihood Ratio Test		
Model Terms	**IRR (95% CI)**	**SE**	**p-value**	**df**	**χ^2 Statistic**	**p-value**
Age						
age1	1.39 (1.14, 1.70)	0.102	0.001	1	9.677	0.002
age2	0.72 (0.67, 0.78)	0.037	0.000	1	54.085	0.000
age3	0.87 (0.67, 1.11)	0.070	0.262	1	1.024	0.312
age4	3.72 (0.81, 17.04)	0.777	0.091	1	1.137	0.286
Water Contact Site						
Akpeakpe	1					
Tutukope	0.25 (0.12, 0.50)	0.357	0.000	6	28.786	0.000
Agbesi	0.28 (0.06, 1.27)	0.769	0.098			
Alabo	0.03 (0.00, 0.29)	1.245	0.003			
Foekope	0.01 (0.00, 0.13)	1.525	0.001			
Sefornu	0.02 (0.00, 0.14)	1.115	0.000			
Abor	0.08 (0.00, 6.47)	2.240	0.260			
Sex						
Female	1					
Male	2.59 (1.33, 5.04)	0.340	0.005	1	8.041	0.005
Frequency of Contact						
Daily	1					
Thrice per week	2.54 (0.38, 17.09)	0.973	0.338	1	0.963	0.326
Treatment in the Past Year						
No	1					
Yes	0.91 (0.06, 13.29)	1.370	0.944	1	0.437	0.509

IRR* : Incidence Rate Ratio obtained by exponentiation of the coefficients.

Table 59: Maximal model for the mean of the binomial process i.e. the probability that the observed outcome in a given individual inhabitant was a false zero as against any other observed outcomes.

	Binary Model			Likelihood Ratio Test		
Model Terms	**IRR (95% CI)**	**SE**	**p-value**	**df**	**χ^2 Statistic**	**p-value**
Age						
age3	0.99 (0.65, 1.49)	0.210	0.952	1	0.004	0.951
age4	2.03 (0.32, 13.00)	0.946	0.453	1	0.734	0.392
Treatment in the Past Year						
No	1					
Yes	85.19 (0.05, 1.5 x 10^4)	3.831	0.246	1	0.120	0.729

Table 60: Provisional model for the mean burden of *S. haematobium* infections harboured per individual participant in our study community.

	Count Model		
Model Terms	**IRR (95% CI)**	**SE**	**p-value**
Age			
age1	1.34 (1.11, 1.61)	0.095	0.002
age2	0.73 (0.69, 0.78)	0.034	0.000
age3	1.10 (1.05, 1.16)	0.025	0.000
Water Contact Site			
Akpeakpe	1		
Tutukope	0.23 (0.11, 0.46)	0.354	0.000
Agbesi	0.29 (0.06, 1.36)	0.790	0.117
Alabo	0.02 (0.00, 0.22)	1.206	0.001
Foekope	0.01 (0.00, 0.16)	1.268	0.001
Sefornu	0.02 (0.00, 0.15)	1.116	0.000
Abor	0.08 (0.00, 6.47)	2.244	0.259
Sex			
Female	1		
Male	2.30 (1.21, 4.36)	0.327	0.011

Table 61: Provisional model for the mean of the binomial process i.e. the probability that the observed outcome in a given individual inhabitant was a false zero as against any other observed outcomes.

	Binary Model		
Model Terms	**IRR (95% CI)**	**SE**	**p-value**
Age			
age3	1.12 (1.04, 1.21)	0.040	0.005

3.2.3.1.2 Selection of the Optimal Model

Using the results of the likelihood ratio tests as described in section 2.5.1.6, a backwards selection by the sequential omission of the least significant terms was performed. Therefore, individual treatment history and frequency of water contact were excluded from the count model, Tables 58, whilst individual treatment history and age4, were excluded from the binomial model (Table 59). Though initially non-significant, the effect of age3 assumed statistical significance in the binomial model after the other non-significant terms were sequentially omitted. Tables 60 and 61, therefore, became our provisional models.

3.2.3.2 Goodness-of-Fit Check

Our assessment of the provisional model for over-dispersion yielded a value of 1.18 for the Pearson dispersion statistic. Since this value fell within the acceptable cut-off range, as discussed in section 2.5.1.7.1, we ruled out any additional unexplained variation in our study outcome. Hence, the provisional model was deemed suitable, conditional on validation.

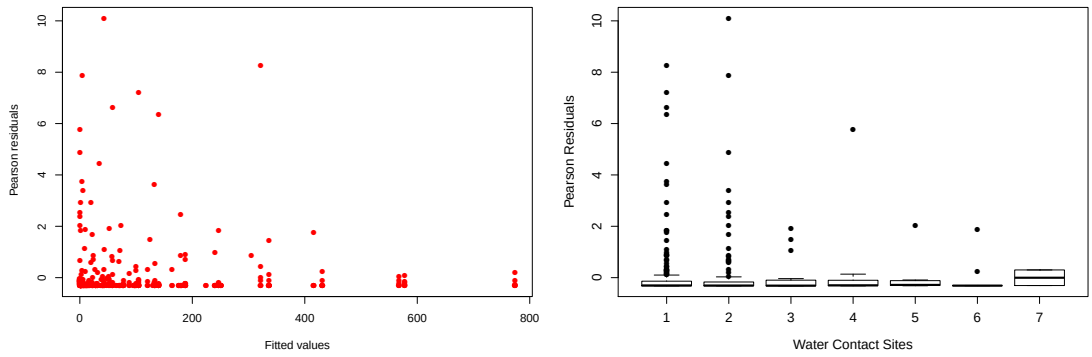
3.2.3.3 Model Validation

3.2.3.3.1 Assessing the Homogeneity Assumption

Rather than making subjective judgement about any observed patterns in Figure 97, we opted for the more unbiased alternative of using generalised additive models with a Gaussian error distribution, equation 22, to assess the effect of each of the continuous-valued covariates in the provisional model on any patterns in the residuals, ϵ_i. Therefore, any relevant residual patterns due to these covariates were expected to manifest as significant smoother(s) (Zuur et al. 2012).

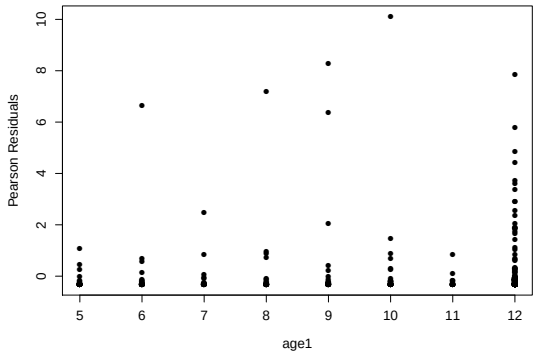
$$\epsilon_i = \alpha + f(\mathrm{X}_{1_i}) + e_i \tag{22}$$

where ϵ_i denotes the Pearson residuals of the provisional model, $f(X_{1_i})$ is the smooth function for a continuous-valued covariate in the provisional model and e_i is the random noise in the data.

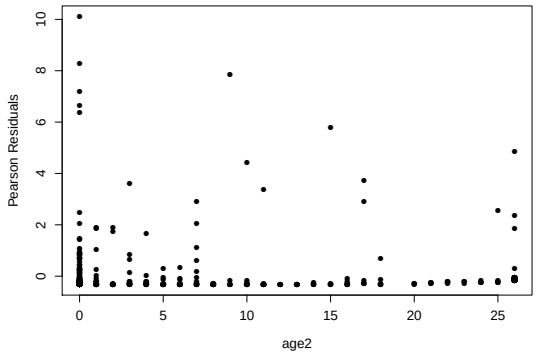


(a)

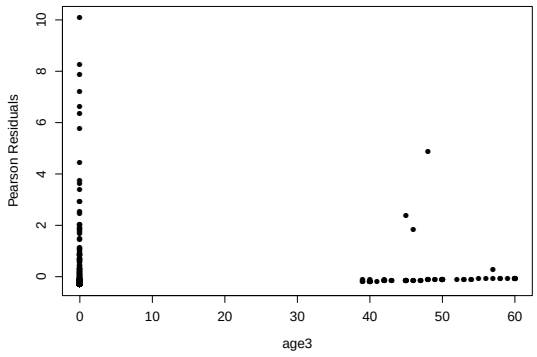
(b)



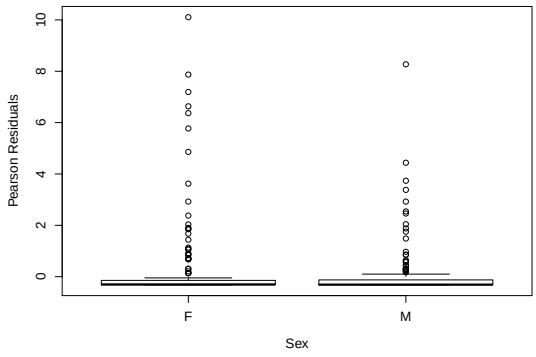
(c)



(d)



(e)



(f)

Figure 97: Assessing the standardised residuals for patterns. A plot of the Pearson residuals of the provisional model versus fitted values (a). Pearson residuals versus individual terms in the provisional model (b - f).

3.2.3.3.2 Assessing Residual Patterns With Generalised Additive Models

The smoothers were used in this case to summarise any trends in the Pearson residuals due to the continuous-valued covariates. Therefore, a significant effect for any of these smoothers would be indicative of significant non-linear trends in the residuals and hence the need to further improve the model. The model improvement in such cases would, therefore, centre around the modelling of the effects of those covariates as non-linear rather than linear terms.

However, judging by the numerical output of the smoothers in Table 62, there was no evidence that any of the continuous-valued covariates had a significant non-linear effect on the residual patterns. Moreover, each of the smoothers barely explained about -0.2 percent of the total variation in the residuals. These results were, therefore, used as basis for ruling out any patterns in the Pearson residuals of the provisional model. We, therefore, concluded that there were no obvious patterns in the residuals that justified any further improvement of our provisional model.

Table 62: Approximate significance of the smooth terms for each of the continuous-valued covariates in the separate Generalised Additive Models that had the Pearson residuals as outcome.

Smooth Terms	EDF	R-Square	Ref. df	F	p-value
s (age1)	1	-0.002	1	0.002	0.961
s (age2)	1	-0.002	1	0.001	0.975
s (age3)	1	-0.002	1	0.000	0.995

3.2.3.3.3 Assessing the Assumption of Independence

3.2.3.3.4 Spatial Autocorrelation

As an initial assessment of the effect of the unmeasured factors, Figure 98 was examined for apparent clusters of positive or negative residuals which might signify the presence of

residual spatial effects. However, although some apparent clustering of negative residuals were mainly obvious in the south-western corner of the community, the distribution of the residuals in the rest of the community was not clearly segregated. A comparison of Figure 98 with the raw data, Figure 99, suggests the model mainly over-estimated the risk of infection in two parts of the community (as indicated by the blue circles in Figure 98).

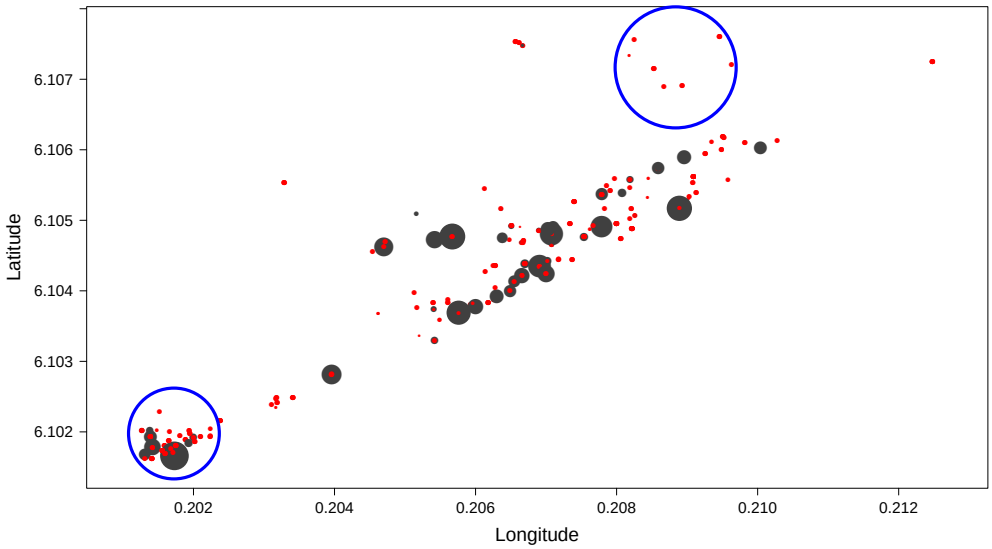


Figure 98: Spatially-referenced residuals of the provisional model for the risk of aggregated *S. haematobium* burden. The radii of the circles are proportional to the absolute values of the Pearson residuals. The red and grey circles represent the negative and positive residuals, respectively.

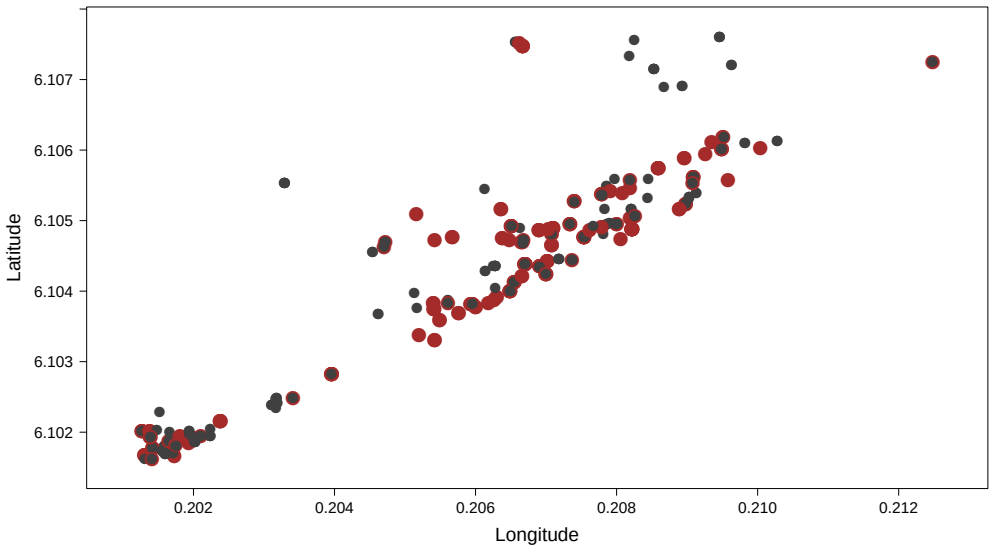


Figure 99: Distribution of the study population in Alabonu by infection status. The position of each point signifies a sampled location.

3.2.3.3.5 Structural Analysis

The empirical variogram, Figure 100, was used as a more objective tool in quantifying any spatial dependence in the residuals. However, while Figure 100 suggests a decreasing spatial trend with lag distance, it is an omni-directional variogram that assumes isotropy. Therefore, to verify this assumption, directional variograms, Figure 101, were computed and examined for differences in their slopes which would point to a violation of the

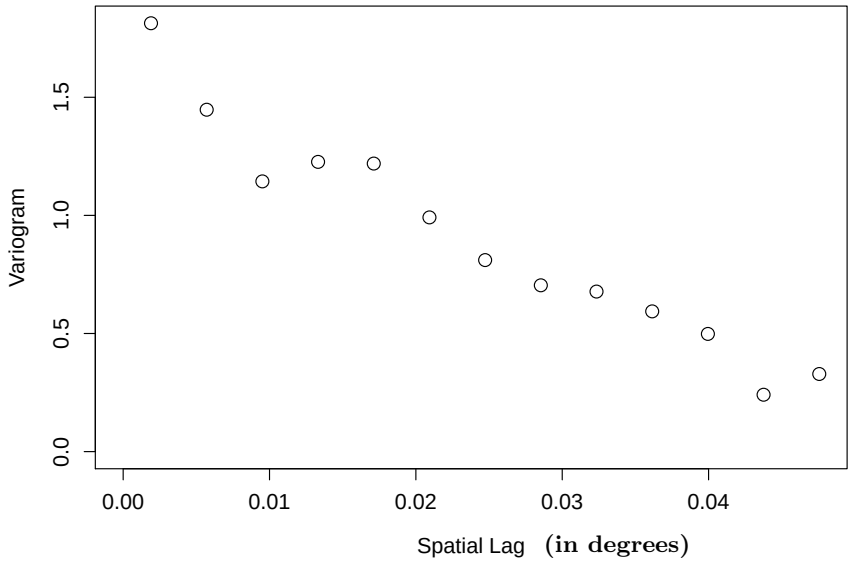


Figure 100: Omni-directional empirical variogram for the Pearson residuals of the provisional model for aggregated *S. haematobium* burden.

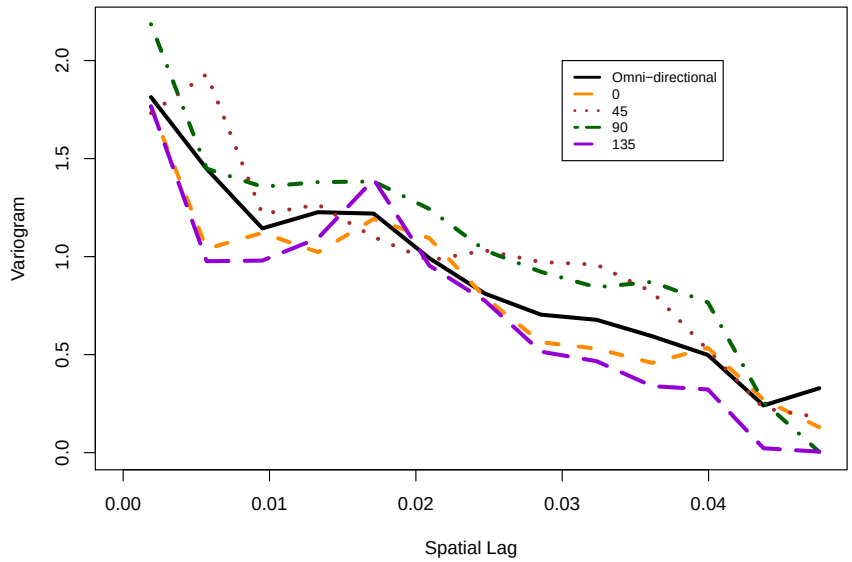


Figure 101: Directional empirical variograms for lags in the 0°, 45°, 90° and 135° spatial directions, with a tolerance angle of ± 22.5°.

isotropy assumption. However, since all the anisotropic empirical variograms in Figure 101 look relatively identical, isotropy seemed like a reasonable assumption. The omni-directional variogram was, therefore, regarded as the mean variogram for all directions and used in the assessment for spatial trend.

However, as judged by the envelope of random permutations for the empirical variogram, Figure 102, and the formal test for trend (p=0.888); there was no evidence of statistically significant spatial autocorrelation. These results were, therefore, used as a basis for ruling out the relevance of the effect of the unmeasured factors on the validity of our provisional model. Hence, the measured covariates seemed adequate in explaining the variation in the risk of aggregated schistosome burden.

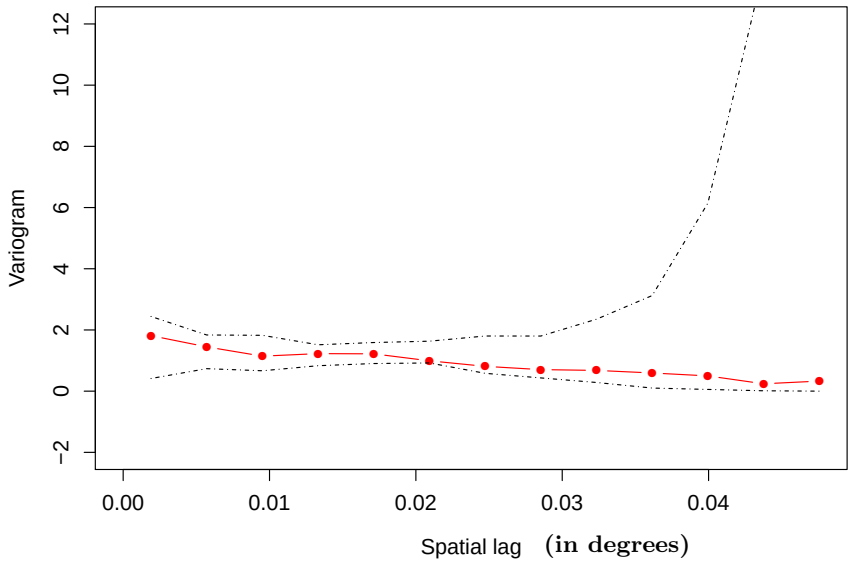


Figure 102: Empirical variogram for the Pearson residuals of the provisional model for aggregated *S. haematobium* burden (red line) and 95% confidence envelope. The variogram falls within the envelope of random permutations, thereby suggesting the absence of any significant spatial dependence in the residuals.

Moreover, the absence of any significant residual spatial effects seemed to rule out our assumption that inhabitants who frequented the same water contact sites were likely to have harboured identical mean burden of infections. Hence, despite the community-level clustering of familial units and the fact that members of the same familial aggregations mostly frequented the same water contact sites, our results seem to suggest varying mean burden of harboured schistosomes among members of the same familial units.

3.2.3.4 Final Model for Aggregated *S. haematobium* Burden

3.2.3.4.1 Count Model

Therefore, having ruled out any significant violation of the homogeneity and independence assumptions, the provisional model, Table 60, became our final model for aggregated schistosome burden in the study community. Hence, holding all other terms constant, the mean burden of schistosomes harboured per individual inhabitant was increased by a factor of 1.34 (95% CI: 1.11 - 1.61) among inhabitants who fell within the age group of age1; and by a factor of 1.10 (95% CI: 1.05 - 1.16) among those in the age3 group. Moreover, compared to females, the male inhabitants had a 2.30 (95% CI: 1.21-4.36) times greater odds of harbouring aggregated schistosome burden.

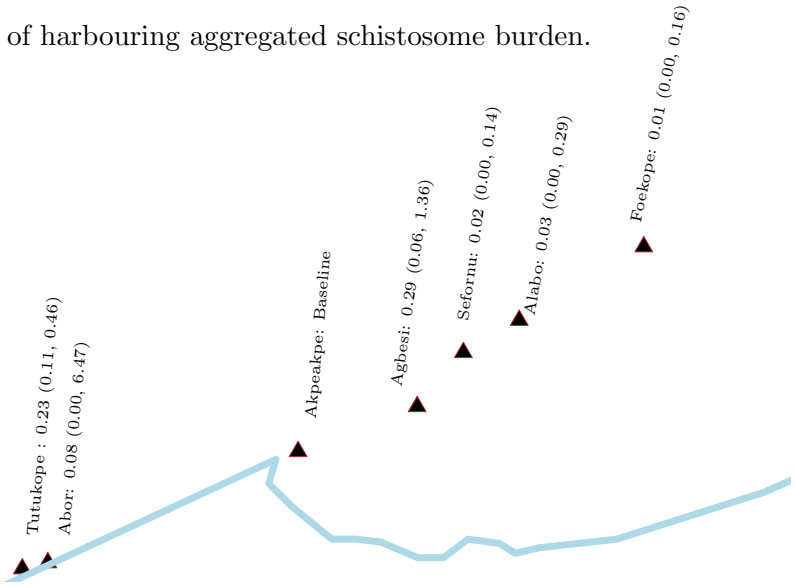


Figure 103: The risk of aggregated schistosome burden associated with frequent exposure to each of the water contact sites in our study community. The incidence rate ratios are presented against each of the sites with the 95% confidence intervals in brackets.

The water contact sites were mainly associated with varying degrees of decreased odds of harbouring aggregated schistosome burden (Figure 103). However, inhabitants who frequently got exposed to the Foekope (0.01, 95% CI: 0.00 - 0.16), Alabo (0.02, 95% CI:0.00 - 0.22) and Sefornu (0.02, 95% CI: 0.00 - 0.15) water contact sites had the lowest odds of harbouring aggregated schistosome burden, as compared to inhabitants who frequented the Akpeakpe water contact site.

3.2.3.4.2 Binary Model

This component of the model, Table 61, describes the change in odds of harbouring false zero outcomes versus all other values of the outcome. The only covariate in the binary model, age3, had an incidence rate ratio of 1.12. This, therefore, implies that despite having an increased odds of harbouring aggregated schistosome burden, as judged by their effect in the count model, inhabitants in that age group were also at a 1.12 (95% CI: 1.04 -1.21) times greater odds of not excreting any *S. haematobium* eggs despite harbouring infections.

4 Discussion

Due to heterogeneities in the transmission of the human schistosome species, heavy intensity infections tend to be focused within a minor fraction of the infected human host population in different endemic settings. These heavily infected individuals are, therefore, responsible for sustaining the transmission of *Schistosoma* infections. While a series of factors including genetics, behavioural and social factors have been implicated in the predisposition to harbouring aggregated schistosome burden (Anderson & Medley 1985); issues relating to the variation in the infectivity of the water contact points where exposure occurs are often overlooked, mainly as a result of the difficulty involved in measuring such factors.

Empirical evidence suggests that human exposure patterns may be distributed across different water contact sites in endemic shoreline settings (Woolhouse et al. 1998, Chandiwana & Woolhouse 1991). Moreover, these water contact sites may vary in their levels of infectivity due to site-specific variations in contamination and pollution rates, and consequently the population density of the intermediate snail hosts harboured by these sites (Woolhouse et al. 1998, Dalton & Pole 1978).

Therefore, the present paper has investigated the variation in the risk of aggregated schistosome burden as a function of factors that are related to both the human host population as well as the intermediate snail host in the water contact sites where exposure occurred. However, factors relating to the varying infectivity of the water contact sites were investigated as missing covariates in our model.

By using data from a pre-intervention shoreline setting, it was possible for us to assume an endemic equilibrium state of transmissions as well as constant exposure patterns across all active age groups in the human hosts population. Moreover, the geostatistical methodologies that were employed in assessing the effect of the unmeasured factors would have enabled us to adapt our model to account for the effect of both the measured and missing covariates, had the effect of the latter been significant.

Our findings suggest that exposure to the various water contact sites in the community was associated with varying odds of harbouring aggregated schistosome burden. Moreover, the male inhabitants as well as inhabitants in the age1 and age3 categories were found to have an increased odds of harbouring aggregated schistosome burden. However, the water contact sites of frequent exposure did not seem to influence the distribution of inhabitants who harboured aggregated schistosome burden in the community. Taken together, these results suggest the unmeasured factors associated with the infectivity of the water contact sites did not induce any significant spatial autocorrelation which would have inflated the model's precision due to the lack of independent information on the autocorrelated residuals. Hence, the effects of the measured covariates in the model were valid.

While the effect of the unmeasured covariates did not seem to influence the community-level variation in the risk of aggregated schistosome burden, we also interpret these results with a bit of caution. This is because the interpretation of the results may be influenced by the fact that regardless of the convergence of the fitting algorithm of the model, the decomposition into large-scale and small-scale variations may not always be reliable. But rather, the reliability of the decomposition may depend on the strength of the covariate effects as well as the amount of variability in the data (Waller & Gotway 2004). Hence, the absence of significant spatial autocorrelation in our case may either signify the lack of any significant effect between the unmeasured factors and our study outcome or just weak covariate effects of the unmeasured factors. It, however, remains to be determined how weak a covariate effect needs to be, scientifically, in order for its effect to appear irrelevant in a model or conversely.

We also interpret our findings in the context of some possible limitations with regard to our data collection method. Firstly, no qualitative studies were conducted to formally assess the exposure patterns of our study population as well as the site-specific variations in exposure activities. Therefore, we mainly relied on interviewer-administered questionnaire responses on the most frequented points of exposure; and some general observations during our field surveys in the community. While these questionnaire responses mainly suggested a concentration of exposure activities at two sites, most of those responses may have been based on canoe boarding points, rather than the specific points where

domestic chores were performed. Therefore, there were virtually no exposure data for some of the community's water contact sites and the analysis had to be restricted to those sites for which we had data.

Our model, however, took the annual cycle of the lake into account. Therefore, these findings are only restricted to the period of stagnation for reasons which we now discuss. Unlike the major rainy season when human exposure patterns were considerably reduced, exposure patterns were highest during the period of stagnation of the lake. Moreover, the virtually stagnant conditions of the water contact sites during that period was likely to confine the effect of pollution by anthropogenic activities specifically to the points of application. Therefore, the resultant effect of the site-specific anthropogenic activities on the ecology of the snail hosts that inhabited those sites were also bound to be at their highest.

The effect of these ecological changes on the snail hosts, be it positive or otherwise, was expected to manifest in the degree of schistosome burden harboured by the different subsets of inhabitants who frequented the various water contact sites. In other words, the cumulative effect of exposure activities that rendered the water contact sites uncongenial for the snail hosts was expected to have induced considerable reductions in their population densities. Consequently, inhabitants who frequented those sites were likely to harbour fewer schistosome infections than those who frequented the sites where exposure activities resulted in ecological changes that favoured the survival of the snail hosts species.

Our assumptions regarding the effect of anthropogenic activities on the ecology of the snail hosts were, however, made without taking the site-specific variation in the abundance of emergent plants into account. Emergent plants such as *Ceratophyllum sp* are known to play crucial roles in the survival of the snail host species. Not only do these plants serve as food and oxygen sources for the matured snails but they also provide substrates for their egg nests (Cowper 1971). The population density of the snail hosts, therefore, tends to vary with the distribution and abundance of these emergent plants (Klumpp & Chu 1980, Scott et al. 1982, Chu et al. 1981).

Therefore, it is possible that the presence of *Ceratophyllum sp* in the water contact sites may have potentially counteracted with the effects of certain anthropogenic activities, including those regarding motor-powered canoes. For instance, interruptions in the air supply of the snail hosts, due to the presence of oil slicks on the surface of the water, may be compensated for by the oxygen supply from the population of *Ceratophyllum sp* in the sites. Moreover, judging by the role of *Ceratophyllum* as substrates for the snails' eggs, it is likely that at least a fraction of the eggs in each site may survive the period of high flow rate during the major rainy season.

Hence, one of our main study assumptions regarding the invasion of the water contact sites by new generations of snail host at the end every major rainy season may not be entirely accurate. But rather, some of the snail host population from previous seasons may be retained through the eggs that remain attached to the population of *Ceratophyllum sp* in the sites. Therefore, the newly hatched population of snails may end up being more resistant to ecological changes, such as those induced by anthropogenic activities, due to their history of previous exposure.

If that were indeed possible, then the population density of these resistant snails may vary across the different water contact sites, depending on the levels of contamination of the sites with human excreta. Empirical evidence suggests the life span of the snail hosts may be influenced by their degree of infection with the schistosome parasites. Therefore, whereas the heavily infected snails tend to die within weeks of infection, the lightly infected ones may live their full span of 3-5 years (Barlow & Muench 1951). The implication of this varying life span of the snail hosts to the present study is that the least contaminated sites may end up harbouring higher densities of resistant snails. Therefore, the interpretation of our model would be incomplete until issues relating to the bionomics of the snail hosts are also taken into account.

5 Conclusions and Recommendations

The effectiveness of schistosomiasis control programmes in endemic settings is considerably improved by the integration of chemotherapy interventions with the necessary operational components of transmission control that result in decreased exposure levels within the human host population. However, the element of decreased exposure has up until this point, virtually remained an unattainable component of control in many endemic settings.

This is mainly due to the fact that the allocation of intervention resources, such as potable water supply, is still very poor in many endemic parts of sub-Saharan Africa as a consequence of the slow pace of socio-economic development. Moreover, the water bodies that serve as sources of transmission tend to double as major sources of livelihood, especially in the endemic shoreline settings. Therefore, it might take several more years, if not decades, for control programmes to successfully integrate all the elements that would eventually make reduced exposure levels a functional component across endemic settings.

All these, therefore, go to show that the effective control of schistosomiasis might in the mean time benefit from alternative strategies that do not rely on the key element of reduction in human exposure levels as a means of sustaining the benefits of the ongoing mass chemotherapy interventions. Therefore, the main challenge centres on making exposure, which is being regarded in this case as an inevitable consequence of living in rural endemic shoreline settings, less infective for snail-to-human transmission of infection. But has that strategy not been tried before? Indeed, it has.

Long before the human definitive hosts became the focus of intervention, early control strategies were mainly directed at the snail host population. Therefore, the focal molluscicide of water contact sites, with the ultimate aim of reducing the infectivity of water bodies that transmitted schistosome infections, were tried and tested across the endemic parts of Africa (Chu et al. 1981, Cowper 1971, Asaolu & Ofoezie 2003). The results of these early trials mainly indicated that synthetic chemicals such as bayluscide niclosamide and sodium pentachlorophenate were highly effective against the interme-

diate snail hosts, but not their eggs (Cowper 1971). Therefore, the application of these chemicals had to be repetitive (Cowper 1971). Moreover, it became evident over time that the decrease in the snail population density, induced by these molluscicides, did not translate into a reduction in the burden of infections within the human host population (Asaolu & Ofoezie 2003).

However, it is possible that the concurrent molluscicide of water contact sites and the administration of chemotherapy in the human host population might fare better in reducing the overall burden of infections which would subsequently translate into a reduction in the transmission potential of infections. This is because the core fraction of heavily infected individuals, who have the greatest tendency of contributing most to environmental contamination, and the specific water contact sites that harbour the majority of infections would be targeted together. Therefore, this strategy, if sustained over time, could eventually lead to an interruption in the transmission of infections (Nasell 1976).

Now, let us assume that the focal molluscicide of water contact sites could be replaced by an alternative strategy that does not rely on the aforementioned chemicals. Building on this concept, we could assume further that such an alternative strategy is already in place, and indeed has been in place for as long as these endemic shoreline communities have been in existence. Moreover, despite having been in place since time immemorial, the effect of this alternative strategy on the ecology of the snail host species has been gradually increasing in severity over time. To demystify this alternative strategy for anyone who might be wondering, we are referring to the effect of anthropogenic activities in endemic shoreline settings on the ecology of the snail hosts, the main subject of the present study. The next practical question, therefore, might be why the effect of these anthropogenic activities has gone unnoticed over the years. Well, that is probably because we have not really been looking.

Variations in water contact activities across different endemic settings have been documented throughout scientific literature (Bruun & Aagaard-Hansen 2008, Farooq & Mallah 1966, Dalton & Pole 1978). Farooq & Mallah (Farooq & Mallah 1966) however, provided a more in-depth insight into the religious and socio-economic context in which

exposure occurred in communities situated along four major canals in Egypt. The main element in all the aforementioned studies is that the inhabitants of these shoreline settings have consistently stuck by the age-old practice of performing exposure-related activities directly inside the water bodies. Therefore, these activities end up having a direct effect on the ecology of the intermediate snail hosts. However, the severity of the effect of these activities has also been changing over the years. For instance, manually-operated canoes are being upgraded into motor-powered ones whilst competition between production companies has led to an improvement in the chemical constituents of soap.

The prevailing question, therefore, is if these changes to the ecology of the snail hosts are capable of inducing molluscicidal effects in the water contact sites? If that were indeed the case, then is it also possible that the combined effect of these anthropogenic activities and chemotherapy interventions within the human hosts population may eventually cause the burden of infections to fall below the critical break point?

The present study has indirectly assessed the collective effect of anthropogenic activities on the ecology of the snail host species. We, however, consider our findings inconclusive on the basis that our data collection method lacked the integral element of a formal qualitative study on how diffused or concentrated the exposure patterns were, and the site specific variation in exposure activities. The assumptions that went into our model include the following: the endemic equilibrium of transmissions was still unperturbed, as far as the effect of chemotherapy interventions in the human hosts population was concerned; that transmissions were autochthonous; and that burden of infections within the human host population was a direct reflection of the level of infectivity of the water contact sites they mainly frequented.

Future research into this particular area would, therefore, benefit from a more formal assessment of human exposure patterns as well as the variations in exposure activities across the different water contact sites in the community. In that regard, Seto et al. 2007 have suggested that global positioning system monitors, when used in conjunction with questionnaires, may provide more reliable data on human exposure/activity patterns.

The present study was conducted in February - March, which falls within the period of stagnation of the lake. However, a better assessment of the effect of these anthropogenic activities on the ecology of the snail host species could be designed in such a way that the study is repeated at different points within the period of stagnation. That way, the results from the different time points could be compared. An assessment of the effect of the site-specific anthropogenic activities on the population of *Ceratophyllum sp* would also allow a better judgement of whether the snail hosts and the population of *Ceratophyllum* are equally affected by anthropogenic activities.

Another effect that warrants further investigation is the differential risk of aggregated schistosome burden for males and females. Social factors such as gendered tasks are generally known to govern exposure patterns in endemic settings (Bruun & Aagaard-Hansen 2008). Therefore, while exposure activities such as the washing of clothes is predominantly a female task, the operation of canoes is exclusively reserved for males (Farooq & Mallah 1966). These tasks may, therefore, influence the infectivity of the specific sites of exposure as well as the time, degree and frequency of exposure in the female and male inhabitants of any specific endemic shoreline setting. Future studies would, therefore, benefit from a more in-depth analysis that accounts for the specific water contact sites that are mostly frequented by the males and females inhabitants in relation to their social tasks as well as interactions between age and sex.

References

Anderson, R. M. & Medley, G. F. (1985), 'Community control of helminth infections of man by mass and selective chemotherapy', *Parasitology* **90**, 629–660.

Arlinghaus, S. L., ed. (1996), *Practical Handbook of Spatial Statistics*, first edn, CRC Press LLC, chapter 8, pp. 175–209.

Asaolu, S. & Ofoezie, I. (2003), 'The role of health education and sanitation in the control of helminth infections', *Acta Tropica* **86**(2-3), 283 – 294.

Barlow, C. H. & Muench, H. (1951), 'Life span and monthly mortality rate of *Bulinus truncatus* and *Planorbis boissyi*, the intermediate hosts of schistosomiasis in Egypt', *The Journal of Parasitology* **37**(2), 165–173.

Basanez, M.-G., French, M., Walker, M. & Churcher, T. (2012), 'Paradigm lost: how parasite control may alter pattern and process in human helminthiases', *Trends in Parasitology* **28**(4), 161 – 171.

Bivand, R. & Rundel, C. (2013), *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. R package version 0.2-12.

Bruun, B. & Aagaard-Hansen, J. (2008), *The social context of schistosomiasis and its control: An introduction and annotated bibliography*, World Health Organization.

Chandiwana, S. K. & Woolhouse, M. E. J. (1991), 'Heterogeneities in water contact patterns and the epidemiology of *Schistosoma haematobium*', *Parasitology* **103**, 363–370.

Cheever, A., Kamel, I., Elwi, A., Mosimann, J. & Danner, R. (1977), '*Schistosoma mansoni* and *S. haematobium* infections in Egypt. II. Quantitative parasitological findings at necropsy', *The American Journal of Tropical Medicine and Hygiene* **26**(4), 702–16.

Chu, K. & Vanderburg, J. (1976), 'Techniques for estimating densities of *Bulinus truncatus rohlfsi* and its horizontal distribution in the Volta Lake, Ghana', *Bulletin of the World Health Organization* **54**, 411–416.

Chu, K. Y., Vanderburg, J. A. & Klumpp, R. K. (1981), 'Transmission dynamics of miracidia of *Schistosoma haematobium* in the Volta Lake', *Bulletin of the World Health Organisation* **59**(4), 555–560.

Cowper, S. (1971), *A Synopsis of African Bilharziasis*, H. K. Lewis.

Crawley, M. J. (2014), *Statistics: An introduction using R*, 2nd edition edn, Wiley-Blackwell.

Dalton, P. & Pole, D. (1978), 'Water-contact patterns in relation to *Schistosoma haematobium* infection', *Bulletin of the World Health Organization* **56**(3), 417–426.

Diggle, P. J. & Ribeiro, P. J. (2007), *Model-based Geostatistics*, Springer.

Eagle, S. & Diggle, P. J. (2012), Spatial independence test for geostatistical data, Technical report, University of Liverpool.

Farooq, M. & Mallah, M. (1966), 'The behavioural pattern of social and religious water-contact activities in Egypt-49 bilharziasis project area', *Bulletin of the World Health Organization* **35**, 377–387.

Gelfand, A., Diggle, P., Guttorp, P. & Fuentes, M. (2010), *Handbook of Spatial Statistics*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, Taylor & Francis.

Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006), 'Human schistosomiasis', *Lancet* **368**(9541), 1106–1118.

Guelat, J. (2013), 'Spatial autocorrelation (modelling)', `http://rstudio-pubs-static.s3.amazonaws.com/9687_cc323b60e5d542449563ff1142163f05.html`.

Gurarie, D. & King, C. H. (2005), 'Heterogeneous model of schistosomiasis transmission and long-term control: the combined influence of spatial variation and age-dependent factors on optimal allocation of drug therapy', *Parasitology* **130**, 49–65.

Guyatt, H. L., Smith, T., Gryseels, B., Lengeler, C., Mshinda, H., Siziya, S., Salanave, B., Mohome, N., Makwala, J., Ngimbi, K. P. & Tanner, M. (1994), 'Aggregation in schistosomiasis: comparison of the relationships between prevalence and intensity in different endemic areas', *Parasitology* **109**, 45–55.

Hilbe, J. (2009), *Logistic Regression Models*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Hilbe, J. M. (2011), *Negative Binomial Regression*, 2nd edn, Cambridge University Press.

Hilbe, J. M. (2014), *Modeling Count Data*, Cambridge University Press.

Jamison, D., Breman, J. & Measham, A., eds (2006), *Disease Control Priorities in Developing Countries.*, 2nd edn, World Bank, pp. 467–481.

Jordan, P. & Webbe, G. (1982), *Schistosomiasis: epidemiology, treatment, and control*, Heinemann.

Jr, P. J. R. & Diggle, P. J. (2016), *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.2.

Keitt, T. H., Bivand, R., Pebesma, E. & Rowlingson, B. (2013), *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 0.8-4.

Kloos, H., Gazzinelli, A. & van Zuyle, P. (1998), 'Microgeographical patterns of schistosomiasis and water contact behaviour; examples from Africa and Brazil', *Mem Inst Oswaldo Cruz* **93**(Suppl. 1), 37–50.

Klumpp, R. K. & Chu, K. (1980), 'Importance of the aquatic weed *Ceratophyllum* to transmission of *Schistosoma haematobium* in the Volta Lake, Ghana', *Bulletin of the World Health Organization* **58**(5), 791–798.

Lustigman, S., Prichard, R., Gazzinelli, A., Grant, W., Boatin, B., McCarthy, J. & Basanez, M. (2012), 'A research agenda for helminth diseases of humans: the problem of helminthiases.', *PLoS Neglected Tropical Diseases* **6**(4), e1582.

Macdonald, G. (1965), 'The dynamics of helminth infections, with special reference to schistosomes', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **59**(5), 489 – 506.

Medley, G. F. & Bundy, D. (1996), 'Dynamic modeling of epidemiologic ppattern of schistosomiasis morbidity', *American Journal of Tropical Medicine and Hygiene* **55**(5), 149–158.

MoFA (2015), 'Sustainable development of rain-fed lowland rice production project', http://mofa.gov.gh/site/?page_id=4632. Online; accessed 19-June-2015.

Nasell, I. (1976), 'On eradication of schistosomiasis', *Theoretical Population Biology* **10**(2), 133 – 144.

Okwuosa, V. & Osuala, F. (1993), 'Toxicity of washing soaps to *Schistosoma mansoni* cercariae and effects of sublethal concentrations on infectivity in mice', *Applied Parasitology* **34**(1), 69–75.

Pearce, E. J. & MacDonald, A. S. (2002), 'The immunobiology of schistosomiasis.', *Nature Reviews Immunology* **2**(7), 499.

Rigby, R. A. & Stasinopoulos, D. M. (2005), 'Generalized additive models for location, scale and shape,(with discussion)', *Applied Statistics* **54**, 507–554.

SCI (2013), 'Schistosomiasis consortium for operational research and evaluation', http://score.uga.edu/SCORE_Brochure_Public_Compressed.pdf.

Scott, D., Senker, K. & England, E. (1982), 'Epidemiology of human *Schistosoma haematobium* infection around Volta Lake, Ghana, 1973-75', *Bulletin of the World Health Organization* **60**(1), 89–100.

Seto, E. Y., Knapp, F., Zhong, B. & Yang, C. (2007), 'The use of a vest equipped with global positioning system to assess water-contact patterns associated with schistosomiasis', *Geospatial Health* **2**, 233–241.

Venables, W. N. & Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.

Waller, L. A. & Gotway, C. A. (2004), *Spatial Exposure Data*, John Wiley & Sons, Inc., chapter 8, pp. 272–324.

Webster, R. & Oliver, M. (2001), *Geostatistics for environmental scientists*, Statistics in practice, John Wiley & Sons.

Weller, Z. (2016), *spTest: Nonparametric Hypothesis Tests of Isotropy and Symmetry*. R package version 0.2.4.

White, P., Webber, D. J. & Helvin, A. (2014), *Modelling structural change using broken sticks*, University of the West of England, Bristol, UK.

Wikipedia (2015), 'Akosombo Dam', `http://en.wikipedia.org/wiki/Akosombo_Dam`. [Online; accessed 01-April-2015].

Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC.

Woolhouse, M. J., Etard, J.-F., Dietz, K., Ndhlovu, P. D. & Chandiwana, S. K. (1998), 'Heterogeneities in schistosome transmission dynamics and control', *Parasitology* **117**, 475–482.

Zeileis, A., Kleiber, C. & Jackman, S. (2008), 'Regression models for count data in R', *Journal of Statistical Software* **27**(8).

Zuur, A. F., Saveliev, A. A. & Ieno, E. N. (2012), *Zero Inflated Models and Generalized Linear Mixed Models with R*, Highland Statistics Ltd.

Zuur, A., Hilbe, J. & Ieno, E. (2013), *A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists*, A Beginner's Guide to, Highland Statistics.

Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. (2011), *Mixed Effects Models and Extensions in Ecology with R.*, Springer.

Overall Conclusion

Conclusion

Taken together, the studies presented in this thesis have mainly highlighted the effect of some of the important covariates that explain the local-level variation in the patterns of schistosome transmission in typical endemic settings. Our modelling strategy adopted an approach that firstly focused on the fitting of classical statistical models. The variograms of the standardised residuals of the preliminary models were then employed in the assessment for spatial dependence and the description of any such dependence with the appropriate correlation structure, where necessary. Therefore, depending on the statistical significance of the dependency, the models were refitted to incorporate the spatial correlation into the covariance matrix. By this approach, the effects of the common factors that influenced transmission at neighbouring locations, due to the focality in schistosome transmission patterns, were taken into account. Therefore, it would seem our final models provided more reliable inferences about the covariate effects, as opposed to classical models.

However, our attempt to identify and disentangle the collective effects of particular setting-specific micro-level determinants of infectivity as spatio-temporal trends, which formed the common theme in this thesis, largely proved ineffective. But rather than interpreting our inability to detect to these spatio-temporal trends as an indication that the micro-level factors were not relevant in explaining the variation in schistosome transmission patterns, we attribute the lack of effect to limitations in our data. First all, the use of parasitological detection in the diagnostic assessment of infection was potentially limiting.

This is mainly due to the fact that parasitological detection methods rely on excreted schistosome eggs as proxy measures for infections. In addition to missing pre-patent infections, these techniques are subject to heterogeneities in egg excretion rates, quality of assessment, sampling limitations and poor sensitivity for low intensity infections. Hence, infections in the older age-groups as well as low transmission and post-intervention sites are often under-estimated by these methods. Moreover, judging by the numerous variations that characterise egg excretion and its measurement by these detection techniques, any observed heterogeneities in the distribution of schistosome infections may be a dir-

ect consequence of the variation in egg output, rather than spatio-temporal patterns in transmission. Therefore, our assessment of spatio-temporal trends may have been confounded by the method of diagnostic assessment. Despite these limitation, however, the observed age-infection patterns in our pre-intervention sites were consistent with other studies where peak infection rates were observed in young adolescents (Colley et al. 2014).

Moreover, while the transmission success of infections may depend largely on the survival and fecundity of the adult worms in the human host, it is the distribution of the intermediate molluscan hosts that determines the zones of transmission (Brooker 2007, Stothard 2009). Hence, it is common for infections to be encountered in places other than where the infected cases occur. Therefore, for a proper assessment of the spatio-temporal patterns of transmission, our analysis could have benefited from data on the intermediate molluscan hosts as well as other physical factors that influence transmission.

Such data on the intermediate hosts are even more important for the Volta Lake due to the massive environmental changes that occurred as a result of its creation. Early studies, following the creation of the Lake, identified *Bulinus truncatus rohlfsi* as the most common intermediate host species (Hunter & Organization 1993). Given the various vegetation in the aquatic ecosystem of these snails that serve as habitats, they are able to survive the annual fluctuations in the water level. However, the ecological changes resulting from human activities in the habitats of these snails may also influence their distribution and ability to transmit infections. Surveys on these snails in the Volta Lake are, therefore, necessary for their distribution to be linked to the patterns of schistosome transmission.

Children in typical endemic settings usually become infected by the age of two. With constant exposure to the sources of infection, the schistosome burden in these children would normally increase over the next 10 year period as their bodies become colonised by the schistosomes (Colley et al. 2014). Therefore, for a proper assessment of the spatio-temporal trends in transmission, future studies could employ a longitudinal design, with a follow-up period of about five years, that focuses on children aged between 2-10 years. Infections could be assessed during the annual cycles of the Lake using Cir-

culating Cathodic Antigen dipstick, supplement by indirect markers such as the presence of haematuria. To allow for proper comparison, four measurements could be taken per annum, distributed across the Lake's cycle. By using this approach, the study would observed incidence cases in these children and the progression of the burden of infection, assuming every infected child would remain infected. Moreover, exposure patterns in this children for could be assessed by adopting Seto et al. 2007's approach of using global positioning system monitors. The suitability of such monitors for a study involving children could, however, prove tricky.

Our inability to detect the spatio-temporal patterns of transmission, which we set out to investigate, goes to emphasise the difficulty involved in building spatially-explicit models for a parasitic infection with one of the most complicated life-cycles. Such studies require detailed designs that take the various sources of uncertainties including, the collection of field data, method of diagnostic assessment and analysis into consideration.

References

Basanez, M.-G., McCarthy, J., French, M., Yang, G.-J., Walker, M., Gambhir, M., Prichard, R. & Churcher, T. (2012), 'A reserach agenda for helminth diseases of humans: modelling for control and elimination', *PLoS Neglected Tropical Diseases* **6**(4), e1548.

Bergquist, R. (2013), 'Good things are worth waiting for', *The American Journal of Tropical Medicine and Hygiene* **88**(3), 409 – 410.

Brooker, S. (2007), 'Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**(1), 1–8.

Brown, D. (2002), *Freshwater Snails Of Africa And Their Medical Importance*, Taylor & Francis.

Cavalcanti, M. G., Silva, L. F., Peralta, R. H., Barreto, M. G. & Peralta, J. M. (2013), 'Schistosomiasis in areas of low endemicity: a new era in diagnosis', *Trends in Parasitology* **29**(2), 75 – 82.

Chu, K., Klumpp, R. & Kofi, D. (1981), 'Results of three years of cercarial transmission control in the Volta Lake', *Bulletin of the World Health Organization* **4**, 549–554.

Colley, D. G., Binder, S., Campbell, C., King, C. H. & Tchuem Tchuenté, L.-A. e. a. (2013), 'A five-country evaluation of a point-of-care circulating cathodic antigen urine assay for the prevalence of schistosoma mansoni', *The American Journal of Tropical Medicine and Hygiene* **88**(3), 426 – 432.

Colley, D. G., Bustinduy, A. L., Secor, W. E. & King, C. H. (2014), 'Human schistosomiasis', *The Lancet* **383**(9936), 2253 – 2264.

Colley, D. G. & Secor, W. E. (2014), 'Immunology of human schistosomiasis', *Parasite Immunology* **36**(8).

Corstjens, P. L. A. M., van Lieshout, L., Zuiderwijk, M., Kornelis, D., Tanke, H. J., Deelder, A. M. & van Dam, G. J. (2008), 'Up-converting phosphor technology-based

lateral flow assay for detection of schistosoma circulating anodic antigen in serum', *American Society of Microbiology* **46**(1), 171–176.

Feldmeier, H. & Poggensee, G. (1993), 'Diagnostic techniques in schistosomiasis control. a review', *Acta Tropica* **52**(4), 205 – 220.

GLOFAL (2006), 'Global View of Food Allergy', `http://www.glofal.org/about.html`.

Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006), 'Human schistosomiasis', *The Lancet* **368**(9541), 1106 – 1118.

Hunter, J. M. (2003), 'Inherited burden of disease: agricultural dams and the persistence of bloody urine (Schistosomiasis hematobium) in the Upper East Region of Ghana, 1959 - 1997', *Social Science & Medicine* **56**(2), 219 – 234.

Hunter, J. M., Rey, L. & Scott, D. (1982), 'Man-made lakes and man-made diseases: Towards a policy resolution', *Social Science & Medicine* **16**(11), 1127–1145.

Hunter, J. & Organization, W. H. (1993), *Parasitic Diseases in Water Resources Development: The Need for Intersectoral Negotiation*, Nonserial Publication, World Health Organization.

Jordan, P. & Webbe, G. (1982), *Schistosomiasis: epidemiology, treatment, and control*, Heinemann.

Lengeler, C., Utzinger, J. A. & Tanner, M. (2002), 'Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa', *Bulletin of the World Health Organization* **80**, 235 – 242.

Mahmoud, A. (2001), *Schistosomiasis*, Tropical Medicine: Science and Practice, Imperial College Press.

Medley, G. & Bundy, D. (1996), 'Dynamic modeling of epidemiologic patterns of schistosomiasis morbidity', *American Journal of Tropical Medicine and Hygiene* **55**(5), 149–158.

Pearce, E. J. & MacDonald, A. S. (2002), 'The immunobiology of schistosomiasis.', *Nature Reviews Immunology* **2**(7), 499.

Rabello, A. (1997), 'Diagnosing schistosomiasis', *Mem Inst Oswaldo Cruz, Rio de Janeiro (Memórias do Instituto Oswaldo Cruz)* **92**(5), 669–676.

Schistoinir (2012), 'Innate immune responses and immunoregulation in schistosomiasis: novel mechanisms in the control of infection and disease', `http://www.york.ac.uk/res/schistoinir/index.html`.

Seto, E. Y., Knapp, F., Zhong, B. & Yang, C. (2007), 'The use of a vest equipped with global positioning system to assess water-contact patterns associated with schistosomiasis', *Geospatial Health* **2**, 233–241.

Stothard, J. R. (2009), 'Improving control of African schistosomiasis: towards effective use of rapid diagnostic tests within an appropriate disease surveillance model', *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**(4), 325 – 332.

World Health Organization (2012), 'Soil-transmitted helminthiases:eliminating soil-transmitted helminthiases as public health problem in children. Progress report 2001-2010 and strategic plan 2011-2020', `http://www.who.int/iris/handle/10665/44804`.

World Health Organization (2013), 'Schistosomiasis: progress report 2001 - 2011 and strategic plan 2012 - 2020', `http://www.who.int/iris/handle/10665/78074`.

Appendices

Part A: Personal Information

ID: _____ **Date:** ____ / ____ / ____

Name: _____

Age: _____ **Sex:** □ **M** □ **F**

Village: _____ **House No:** _____

Part B: Contact With the Lake

1. **What is your main source of water supply for domestic use?** _____

2. **Do you come into contact with the Lake or any other water body in your day-to-day activities?**
   - □ Yes, daily
   - □ Yes, 1-3 times per week
   - □ Yes, 4-6 times per week
   - □ Yes, once a month
   - □ Yes, once a year
   - □ Never

3. **What activities/reasons bring you into contact with the Lake? (Tick all that apply)**
   - □ Collecting water for household chores
   - □ Wading through on the way to the farm
   - □ Washing (of cooking utensils, clothes, etc)
   - □ Swimming
   - □ Fishing
   - □ Boarding of canoes
   - □ Sand winning from the river bed
   - □ Other (please specify): _____

Part C: Treatment History for Soil-Transmitted Helminths

4. **Have you ever been de-wormed?**
   - □ Yes, once □, twice □, three times or more □
   - □ Never
   - □ Don't know

5. If answer to Question 4 is "yes", when was the last time you got de-wormed?

- ☐ Up to 1 year ago
- ☐ Up to 2 years ago
- ☐ Up to 3 years ago
- ☐ More than 3 years ago

6. Who administered the de-worming and where? (Tick all that apply)

- ☐ Dispensary staff at a hospital/health centre
- ☐ Staff from hospital/health centre in community
- ☐ Volta River Authority staff
- ☐ The de-wormer was bought by participants (or parents)
- ☐ At school

Part D: Schistosomiasis-Related Information

7. Do you know about schistosomiasis?

- ☐ Yes, very well
- ☐ Yes, but not very well
- ☐ No

The following are symptoms of schistosomiasis: True/False/Don't know

8. Passing blood in urine?

- ☐ True
- ☐ Flase
- ☐ Don't know

9. Feeling pains when urinating?

- ☐ True
- ☐ Flase
- ☐ Don't know

10. Passing blood in stool?

- ☐ True
- ☐ Flase
- ☐ Don't know

11. Do you pass blood in your urine?

- ☐ Yes
- ☐ No

12. Do you feel pains when urinating?

- □ Yes
- □ No

13. Do you pass blood in your stool?

- □ Yes
- □ No

14. For respondents who answered "yes" to the questions 11-13: What did you do about it?

- □ I visited a hospital/health centre
- □ I used herbal medicine
- □ I (or my parents) bought medicine from a drug store
- □ I did not do anything about it

15. Have you ever been treated for schistosomiasis?

- □ Yes, once □, twice □, three times or more □
- □ Never
- □ Don't know

16. If respondent answered "yes", to Question 15: When was the last time you received treatment?

- □ Up to 1 year ago
- □ Up to 2 years ago
- □ Up to 3 years ago
- □ More than 3 years ago

17. If respondent answered "yes", to Question 15: Who administered the treatment and where?

- □ Dispensary staff at a hospital/health centre
- □ Staff from hospital/health centre in the community
- □ Volta River Authority staff
- □ Bought medicine
- □ At school

Appendix B

Simple logistic regression models for *S. haematobium* infections in Klamadaboe.

		Simple Regression Models			Deviance Analysis		
Model Terms	**No. of Participants**	**OR (95% CI)**	**SE**	**p-value**	D_0 - D_1	*df*	**p-value**
History of Praziquantel Treatment							
Once	19	1					
Twice or more	4	0.00 (NA, 5.24)	2284.102	0.994			
Never	87	0.17 (0.03, 1.00)	0.866	0.041	4.196	2	0.123
Schistosomiasis Awareness							
Yes, very well	16	1					
Yes, but not very well	51	7.71×10^6 (0, NA)	1743.250	0.993			
No	47	7.89×10^6 (0, NA)	1743.250	0.993	1.701	2	0.427
Sex							
Females	69	1					
Males	66	2.46 (0.65, 11.82)	0.714	0.208	1.715	1	0.190
Frequency of Contact with the Lake							
1-3 times per week	7	1					
1 per month	0						
Daily	107	0.31 (0.04, 6.41)	1.173	0.317	0.811	1	0.368
Last Praziquantel Treatment							
Up to a year ago	4	1					
Over 2 years ago	19	0.27 (0.02, 7.25)	1.438	0.358	0.774	1	0.379
Source of Water for Domestic Use							
Tap	2	1					
Lake	107	2.19×10^6 (0, NA)	2797.000	0.996			
Both tap and Lake	1	1 (0, 1.20×10^{115})	4845.000	1.000	0.297	2	0.862
Water Collection							
No	24	1					
Yes	90	0.24 (0.04, 1.38)	0.8536	0.096	2.619	1	0.106
Swimming							
No	49	1					
Yes	65	3.88 (0.60, 75.72)	1.113	0.223	1.894	1	0.169
Sand Winning							
No	98						
Yes	16	0 (NA, 1.48×10^{72})	1809.055	0.993	1.571	1	0.210
Fishing							
No	100	1					
Yes	14	0 (NA, 2.60×10^{75})	1882.924	0.993	1.443	1	0.230
Canoe Boarding							
No	70	1					
Yes	44	0.30 (0.02, 1.96)	1.114	0.283	1.438	1	0.230
Fording							
No	110	1					
Yes	4	0 (NA, 2.03×10^{91})	1978.090	0.994	0.462	1	0.497
Washing							
No	63	1					
Yes	51	0.62 (0.08, 3.32)	0.888	0.590	0.303	1	0.582

Date: ____ / ____ / ____ **Country:** GHANA

Name of interviewer: _____

Child's Details

1. **Name / ID number**: _____ _____
2. **Date of birth / Age**: ____ / ____ / ____ _____ years
3. **Place of birth:** _____
4. **Sex:** □ Male □ Female

5. **School Information**

   Class: _____

   Name of school: _____

6. **House Information**

   House Number: _____

   Suburb/Area: _____

7. **Global Positioning System (GPS) Readings of Child's Home**

   Latitude: _____

   Longitude: _____

   Altitude: _____

8. **How does the child get to school most of the time?**
   - □ walk
   - □ bicycle
   - □ taxi
   - □ bus/trotro
   - □ private car
   - □ Other: _____

Socio-economic Status and Environmental Factors

9. Has the child lived in this town/village since birth? ☐ Yes ☐ No

10. If you answered "No" where has the child lived before and for how long?

Area A In _____ for _____ months _____ years

Area B In _____ for _____ months _____ years

11. Who provides financially for this child?

☐ father and mother

☐ father

☐ mother

☐ Other (please specify): _____

12. Occupation of person in Question 11: _____

Occupation of the spouse of this person: _____

13. The highest level of formal education completed by person in Question 11:

☐ primary/elementary ☐ middle school

☐ junior secondary school ☐ senior secondary school

☐ O-level ☐ A-level

☐ vocational/commercial

☐ training college

☐ polytechnic/university

☐ Other (please specify): _____

14. The house in which the child lives is primarily made of:

☐ cement

☐ wood

☐ mud

☐ Other (please specify): _____

15. What is the main source of water supply in the house:

☐ pipe-borne

☐ tanker

☐ river/stream

☐ well/bore-well

☐ Other (please specify): _____

16. What is the type of toilet in the house:

- ☐ indoor wc
- ☐ compound latrine
- ☐ public neighbourhood latrine
- ☐ Other (please specify): _____

17. The fuel mostly used at home for cooking is:

- ☐ liquified petroleum gas
- ☐ electricity
- ☐ charcoal
- ☐ firewood
- ☐ kerosene
- ☐ Other (please specify): _____

18. What kind of accommodation does child live in?

- ☐ detached house
- ☐ semi-detached house
- ☐ flat
- ☐ compound house
- ☐ Other (please specify): _____

19. Is electricity supplied to your home? ☐ Yes ☐ No

History of Periodic Home-Administered Anthelmintic Treatment

20. When did this child have treatment for worm infection?

- ☐ $<$ 1 month ago
- ☐ 1-3 months ago
- ☐ 3-6 months ago
- ☐ $>$ 6 months ago
- ☐ No idea

21. What is the name of the medicine used in de-worming the child: _____

Appendix D: Copyright Permissions

License Details

This Agreement between Irene Larbi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

printable details

License Number	3751690296500
License date	Nov 17, 2015
Licensed Content Publisher	Elsevier
Licensed Content Publication	The Lancet
Licensed Content Title	Human schistosomiasis
Licensed Content Author	Bruno Gryseels,Katja Polman,Jan Clerinx,Luc Kestens
Licensed Content Date	23–29 September 2006
Licensed Content Volume	368
Licensed Content Issue	9541
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Figure 1
Title of your thesis/dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Estimated size (number of pages)	237
Elsevier VAT number	GB 494 6272 12
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and Nature Publishing Group ("Nature Publishing Group") consists of your license details and the terms and conditions provided by Nature Publishing Group and Copyright Clearance Center.

printable details

License Number	3751690938770
License date	Nov 17, 2015
Licensed Content Publisher	Nature Publishing Group
Licensed Content Publication	Nature Reviews Immunology
Licensed Content Title	The immunobiology of schistosomiasis
Licensed Content Author	Edward J. Pearce, Andrew S. MacDonald
Licensed Content Date	Jul 1, 2002
Licensed Content Volume	2
Licensed Content Issue	7
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1
Author of this NPG article	no
Your reference number	
Title of your thesis / dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Estimated size (number of pages)	237
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

printable details

License Number	3751700424235
License date	Nov 17, 2015
Licensed Content Publisher	Elsevier
Licensed Content Publication	Acta Tropica
Licensed Content Title	The global epidemiological situation of schistosomiasis and new approaches to control and research
Licensed Content Author	D. Engels,L. Chitsulo,A. Montresor,L. Savioli
Licensed Content Date	May 2002
Licensed Content Volume	82
Licensed Content Issue	2
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Fig. 2
Title of your thesis/dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Estimated size (number of pages)	237
Elsevier VAT number	GB 494 6272 12
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

printable details

License Number	3751970363837
License date	Nov 18, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Wiley Books
Licensed Content Title	Applied Spatial Statistics for Public Health Data
Licensed Content Author	Lance A. Waller, Carol A. Gotway
Licensed Content Date	Jul 1, 2004
Licensed Content Pages	520
Type of Use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	2
Original Wiley figure/table number(s)	Fig. 8.1 and Fig. 8.9
Will you be translating?	No
Title of your thesis / dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Expected size (number of pages)	237
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Publisher Tax ID	EU 826 000 141
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

printable details

License Number	3752271269427
License date	Nov 18, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Wiley Books
Licensed Content Title	Geostatistical Error Management: Quantifying Uncertainty for Environmental Sampling and Mapping
Licensed Content Author	Jeffrey C. Myers
Licensed Content Date	Apr 1, 1997
Licensed Content Pages	571
Type of Use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 10-36
Will you be translating?	No
Title of your thesis / dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Expected size (number of pages)	237
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Publisher Tax ID	EU 826 000 141
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and Cambridge University Press ("Cambridge University Press") consists of your license details and the terms and conditions provided by Cambridge University Press and Copyright Clearance Center.

printable details

License Number	3776241486017
License date	Dec 25, 2015
Licensed Content Publisher	Cambridge University Press
Licensed Content Publication	Parasitology
Licensed Content Title	Factors affecting the intensity of reinfection with *Schistosoma haematobium* following treatment with praziquantel
Licensed Content Author	S. K. Chandiwana, M. E. J. Woolhouse and M. Bradley
Licensed Content Date	Apr 6, 2009
Licensed Content Volume	102
Licensed Content Issue	01
Start page	73
End page	83
Type of Use	Dissertation/Thesis
Requestor type	Not-for-profit
Portion	Text extract
Number of pages requested	1
Order reference number	
Territory for reuse	World
Title of your thesis / dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2015
Estimated size(pages)	237
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Billing Type	Invoice
Billing address	Irene Akosua Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, United Kingdom LA1 4YG Attn: Irene Akosua Larbi
Total	**0.00 GBP**
Total	**0.00 GBP**

License Details

This Agreement between Irene Larbi ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

printable details

License Number	3997981037438
License date	Nov 28, 2016
Licensed Content Publisher	Elsevier
Licensed Content Publication	Acta Tropica
Licensed Content Title	Diagnostic techniques in schistosomiasis control. A review
Licensed Content Author	Hermann Feldmeier,Gabriele Poggensee
Licensed Content Date	January 1993
Licensed Content Volume	52
Licensed Content Issue	4
Licensed Content Pages	16
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Order reference number	
Original figure numbers	Table 1
Title of your thesis/dissertation	Social and Environmental Epidemiology of Schistosomiasis in Ghana
Expected completion date	Dec 2016
Estimated size (number of pages)	280
Elsevier VAT number	GB 494 6272 12
Requestor Location	Irene Larbi CHICAS Faculty of Health Sciences Furness College, Lancaster University Lancaster, LA1 4YG United Kingdom Attn: Irene Akosua Larbi
Total	**0.00 GBP**