Spatial epidemiology of parasitic infections and optimal survey design

Hugh Sturrock

Faculty of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine
(University of London)
Thesis submitted for the degree of Doctor of Philosophy (Ph.D.)

April 2011

DECLARATION BY CANDIDATE

I, Hugh Sturrock, confirm that the work presented in this thesis is my own. Where information has

been derived from other sources, I confirm that this has been indicated in the thesis.

Signed..   Date..14th September 2011

HUGH STURROCK

ABSTRACT

Recent years have seen a shift towards integrated control of a number of tropical diseases. Such a strategy, however, relies on an understanding of the spatial distribution and overlap of different diseases. Using a combination of fieldwork, spatial and economic analyses and computerized simulations, optimal survey designs were explored for soil-transmitted helminths (STHs), *Schistosoma mansoni* and *Plasmodium falciparum* in East Africa, and the potential of an integrated survey approach was evaluated.

For STH, analysis indicated that hookworm clusters over larger scales than *Ascaris lumbricoides* and *Trichuris trichiura*, and that surveying small numbers of children, from four to five schools per district, provides a rapid and cost-effective approach to target treatment at district levels. For *S. mansoni*, Lot Quality Assurance Sampling (LQAS) was compared to a geostatistical survey design that allows spatial prediction at unsurveyed locations based on a subset of schools. Results showed that targeted treatment was more cost-effective than presumptive treatment and that, whilst LQAS correctly classified a higher proportion of schools requiring treatment, a geostatistical design proved more cost-effective. An investigation into the optimal spatial scale to conduct surveys for STH, *S. mansoni* and *P. falciparum* in Kenya found that, over various cost scenarios, surveying fifty children from three randomly selected sites per sub-district provided a balance of performance and cost-effectiveness for all species. In sub-districts of low *S. mansoni* and *P. falciparum* prevalence, LQAS should be used to target treatment.

This thesis has shown that species-specific differences in spatial heterogeneity of infection and the costs of both mapping surveys and programme intervention have important implications for the optimal design of surveys. A two stage framework for integrated surveys is proposed allowing for a flexible approach to mapping. Similar studies in different settings are crucial and would help to assess whether changes in survey strategy are required as transmission drops due to control activities.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

ACT	Artemisinin-Combination Therapy
APOC	African Programme for Onchocerciasis Control
CCA	Circulating Cathodic Antigen
CDTI	Community-Directed Treatment with Ivermectin
DALYs	Disability Adjusted Life Year
DEC	Diethylcarbamazine
EIR	Entomological Inoculation Rate
GAHI	Global Atlas of Helminth Infections
GIS	Geographical Information Systems
ICT	Immunochromatographic Card Tests
IPT	Intermittent Preventive Treatment
IQR	Inter-Quartile Range
IRS	Indoor Residual Spraying
ITN	Insecticide Treated Net
IU	Implementation Unit
KEMRI-WTRP	Kenya Medical Research Institute – Wellcome Trust Research Programme
LF	Lymphatic Filariasis
LpCP	Lattice plus Close Pairs
LQAS	Lot Quality Assurance Sampling
MDA	Mass Drug Administration
MIS	Malaria Indicator Survey
MoH-GoSS	Ministry of Health – Government of Southern Sudan
NTDs	Neglected Tropical Diseases
PCT	Preventive Chemotherapy
PDA	Personal Digital Assistant
PR	Parasite Rate
RAGFIL	Rapid Geographical Assessment of Bancroftian Filariasis
RAPLOA	Rapid Assessment Procedure for *Loa loa*
RDT	Rapid Diagnostic Test
REMO	Rapid Epidemiological Mapping of Onchocerciasis
SHNP	School Health and Nutrition Programme
SSA	Sub-Saharan Africa
STH	Soil-Transmitted Helminths
WHO	World Health Organization

CHAPTER 1 | INTRODUCTION

1.1 CONTEXT OF THESIS

In recent years, there has been renewed interest in tropical parasitic diseases and their control. Progress has been made in the control of a range of tropical diseases including malaria (Gething et al., 2010; O'Meara et al., 2010; Snow and Marsh, 2010) and the so-called Neglected Tropical Diseases (NTDs) such as Soil-Transmitted Helminthiasis (STH), schistosomiasis, Lymphatic Filariasis (LF), onchocerciasis and trachoma (Dodd and Cassels, 2006; Burton and Mabey, 2009; Feasey et al., 2009; Hooper et al., 2009; Molyneux et al., 2009; Brooker et al., 2010; Liese et al., 2010). Given similarities in control approaches, and the apparent geographic overlap of these diseases, it has recently been suggested that their control should be integrated on the grounds of cost-savings (Molyneux et al., 2005; Brady et al., 2006; Hotez et al., 2006b; Lammie et al., 2006). This is particularly pertinent for the integration of NTD control as it is estimated that over 90% of NTD burden could be targeted with the administration of just 4 orally administered efficacious drugs: praziquantel; albendazole; azithromycin and ivermectin, that can be safely co-administered (Molyneux et al., 2005; Na-Bangchang et al., 2006; Olsen, 2007). Equally, the possible exacerbating effects of coinfection with NTDs and malaria (Brooker et al., 2006c; Hotez et al., 2006b; Mwangi et al., 2006), make integrated control an appealing approach.

A wealth of theoretical and empirical studies show that control of NTDs and malaria is most cost-effective when targeted at those with the highest risk of morbidity and to the population group who contribute most to transmission dynamics (Anderson and May, 1979; Anderson and May, 1985; Schad and Anderson, 1985; Anderson, 1986; Bundy, 1990; Warren et al., 1993; Smith et al.,

2005). Whilst early studies traditionally focused on the age specificity of infection and its implications for control programmes, it is evident that the spatial dimension of parasite transmission is also important for the targeting of control efforts as the transmission of parasites is spatially heterogeneous (Bundy, 1990; Snow *et al.*, 1996; Brooker and Michael, 2000; Carter *et al.*, 2000; Hay *et al.*, 2006a). This realisation, coupled with technological advances in Geographical Information Systems (GIS) and remote sensing, has given rise to a renaissance in geographic mapping of diseases, including STH (Brooker *et al.*, 2002; Brooker *et al.*, 2009b; Clements *et al.*, 2010a), schistosomiasis (Raso *et al.*, 2005; Clements *et al.*, 2006b; Clements *et al.*, 2008b; Simoonga *et al.*, 2009), LF (Lindsay and Thomas, 2000), trachoma (Polack *et al.*, 2005; Clements *et al.*, 2010b), loiasis (Thomson *et al.*, 2004; Diggle *et al.*, 2007; Crainiceanu *et al.*, 2008), trypanosomiasis (Rogers and Randolph, 1993; Robinson, 1998; Cecchi *et al.*, 2009) and malaria (Kleinschmidt *et al.*, 2001; Hay *et al.*, 2006a; Hay *et al.*, 2009). These maps have, in turn, been useful for identifying priority areas for control and for excluding regions where transmission is absent. However, as control efforts are scaled up, and transmission levels are reduced, there is a scientific need to map infection at the fringes of transmission and to identify remaining hotspots if the goals of control efforts are to be achieved. In the case of STH and schistosomiasis, there is the additional requirement of determining when and where to shift mass treatment from once per year to less frequent intervals (Brooker *et al.*, 2010). Finally, as integrated approaches to the control of diseases becomes more widespread, there is a need to determine the geographic overlap of different diseases.

In order to meet these challenges, information on the distribution and prevalence of infection and disease is required. Traditionally, population-based surveys have been used for this purpose, but

due to the prohibitive financial and technical resource requirements for such surveys, a range of rapid survey methods have recently been developed which aim to provide the minimum amount of information required to make decisions on control (Lengeler *et al.*, 2002; Takougang *et al.*, 2002; Brooker *et al.*, 2005; Brooker *et al.*, 2009a). As a number of diseases display considerable overlap, an intuitively appealing next step is to integrate rapid surveys for individual diseases, taking advantage of shared survey expenditure to maximize cost-effectiveness (Brooker and Utzinger, 2007; Brooker *et al.*, 2009a; Baker *et al.*, 2010). To date, however, there are few guidelines available for an integrated approach to disease mapping, despite recent efforts in the field. In addition, only a few survey methods consider both the spatial characteristics of diseases and the cost of different sampling methods, which both likely have important implications for optimal survey designs. With advances in spatial statistical methods, there is now an opportunity to quantifiably investigate and take into account spatial aspects of diseases, alongside cost, in the evaluation and design of surveys.

This thesis will attempt to rigorously examine survey methods for soil-transmitted helminthiasis, intestinal schistosomiasis, and *Plasmodium falciparum* malaria, accounting for spatial heterogeneity and cost, with the aim of devising an integrated approach to conducting surveys. In addition to varying spatial characteristics, differences in the biology, ecology, epidemiology and diagnostic techniques can influence the design of surveys and the ease with which they can be integrated. This introduction presents an overview of these features, to identify potential challenges and opportunities for integration. The remaining chapters will attempt to address these issues to develop possible approaches to conducting cost-effective integrated surveys.

1.2 THE DISEASES CONSIDERED IN THIS THESIS

This thesis focuses on three parasitic diseases: soil-transmitted helminthiasis (caused by hookworm, *Ascaris lumbricoides* or *Trichuris trichiura*), intestinal schistosomiasis (caused by *Schistosoma mansoni*) and malaria (caused by *Plasmodium falciparum*). These diseases are selected due to the similar targeting of school children during surveys, the potential geographic overlap and the operational need for rapid survey methods to target control according to location. Urinary schistosomiasis due to the trematode *S. haematobium* is not considered here owing to the existence of a reliable, extensively validated, rapid assessment method using school-based questionnaire surveys of reported blood in urine (Lengeler *et al.*, 1991; Lengeler *et al.*, 2002).

1.2.1 Soil-transmitted helminthiasis

Soil-transmitted helminthiasis is caused by infection by any of four STH species: the hookworms (*Ancylostoma duodenale* and *Necator americanus*), roundworm (*A. lumbricoides*) and whipworm (*T. trichiura*). In sub-Saharan Africa (SSA), it is estimated that over 200 million people are infected by at least one STH species (de Silva *et al.*, 2003), with co-infection being common (Booth *et al.*, 1998; Howard *et al.*, 2001). STH infections tend to be characterized by chronic morbidity and disability, rather than mortality. Estimates of Disability Adjusted Life Years (DALYs) due to STH infections vary dramatically, ranging from almost 2.6 million to 39 million (Hotez *et al.*, 2006a; Brooker, 2010), with 2002 World Health Organization (WHO) estimates putting the figure at 6 million (WHO, 2002) (Table 1.1). Infection with *A. lumbricoides* and *T. trichiura* can result in malnutrition, poor growth and cognitive impairment in children (Crompton and Nesheim, 2002; Taylor-Robinson *et al.*, 2007; Hall *et al.*, 2008) and hookworm is associated with iron-deficiency

anaemia (Brooker *et al.*, 2008a; Smith and Brooker, 2010). The mainstay of STH control is periodic,

mass treatment of entire populations with single-dose oral therapy using benzimidazoles,

predominantly albendazole and mebendazole (refer to Table 1.2).

Table 1.1 Disease Burden of malaria and the selected NTDs in Deaths and DALYs (compiled from WHO (2002))

Disease	Deaths	DALYs
Malaria	1,124,000	42.3 million
Schistosomiasis [1]	15,000	1.8 million
Ascariasis	4,000	1.2 million
Trichiuriasis	2,000	1.6 million
Hookworm infection	4,000	1.8 million

[1] Calculated for all major human species: *S. mansoni, S. haematobium* and *S. japonicum*

1.2.2 Intestinal schistosomiasis

Schistosomiasis is caused by infection by worms of the genus *Schistosoma*. The three most

prevalent species are *S. haematobium* and *S. mansoni* which are found in South America, the

Caribbean and Africa and *S. japonicum* which is found in Asia (Chitsulo *et al.*, 2000). Recent

estimates of deaths caused by schistosomiasis do not differentiate between species but

collectively schistosomiasis is estimated to cause 15,000 deaths annually and 1.8 million DALYs,

with the majority of burden occurring in Africa (Chitsulo *et al.*, 2000) (Table 1.1). Infection can

initially result in acute schistosomiasis (Katayama fever), causing fever, headaches and body pain

(Ross *et al.*, 2002). As infection persists, eggs become trapped in tissue and cause formation of granulomatous reactions, which predispose the tissue to malignancy. Most of the pathology is, however, caused by eggs that are washed away by blood and become trapped in small vessels of the liver or lung. (Gryseels *et al.*, 2006). Effective treatment of infection is provided using praziquantel (single oral dose of 40 mg/kg), and WHO currently recommends mass treatment regardless of infection in communities where prevalence exceeds 10% (WHO, 2006c) (Table 1.2). Other control measures include control of the intermediate snail host, improved water and sanitation, and hygienic behaviour through health education.

1.2.3 Malaria

Malaria is a mosquito-borne disease caused by infection with any of five *Plasmodium* spp. parasites: *P. falciparum*; *P. vivax*; *P. malariae*; *P. ovale*; or *P. knowlesi*. Of these five, *P. falciparum* is responsible for the greatest morbidity and mortality, causing approximately 451 million clinical cases (Hay *et al.*, 2011) and around a million deaths per year (WHO, 2002). The majority of mortality from *P. falciparum* occurs in SSA, with around 65% of mortality estimated to occur in children under the age of 5 (Snow and Omumbo, 2006). Treatment is provided by prompt use of antimalarial drugs such as Artemisinin-based Combination Therapies (ACT). A number of other preventative strategies such as distribution of Insecticide Treated Nets (ITNs), Indoor Residual Spraying (IRS), larviciding and Intermittent Preventive Treatment (IPT) are also used. As yet, there are no clear guidelines on which combinations of control methods should be used in a given setting, however, recent work has suggested the use of three endemicity classes: <5%; ≥5 and < 40%; and ≥40% (Hay *et al.*, 2008). These thresholds are based on mathematical models which

predict that in areas where prevalence is ≥40%, the use of bednets alone is not sufficient to interrupt transmission. A lower threshold of 5%, as a point at which surveillance activities should begin, is suggested as an alternative to the 2-3% used by the Global Malaria Eradication Programme (Hay *et al.*, 2008). This revision is again based on evidence from mathematical models which suggests that at these low prevalence levels, transmission is likely to be minimial, yet the sample size requirements to detect such low prevalence are large.

Whilst not a focus of this thesis, the following section gives a brief description of some of the other major NTDs to allow discussion of survey design and control of these diseases in the context of integrated control programmes (see Table 1.2).

1.2.4 Other neglected tropical diseases

LF is a mosquito-borne disease caused by infection with one of three parasitic nematodes: *Wuchereria bancrofti*; *Brugia malayi*; or *Brugia timori*. It has previously been estimated that over 120 million people in 83 countries are infected, of whom 107 million are infected with *W. bancrofti* (Michael and Bundy, 1997), which causes almost 6 million DALYs (WHO, 2008). In SSA, LF is due to *W. bancrofti* and is transmitted by *Culicine*, *Anopheline* and *Aedes* mosquito species. Adult worms live in local lymphatics and release first-stage larvae called microfilariae which can migrate to the blood system and be ingested by mosquitoes during a blood meal. Increased pressure in lymphatics results in lymphoedema and hydrocoele (Kazura, 2002). Control is based on mass treatment of entire populations with recommended anthelmintics: diethylcarbamazine

(DEC) and albendazole outside of SSA. Because of the toxicities of DEC in individuals with onchocerciasis and loiasis, DEC is substituted with ivermectin (WHO, 2000b).

Trachoma is a chronic keratoconjunctivitis caused by repeated reinfection with specific serovars of *Chlamydia trachomatis* (the other serovars cause genital tract disease). It is estimated to cause 84 million cases of active disease worldwide and 1.3 million DALYs (Mecaskey *et al.*, 2003; World Bank, 2004; WHO, 2008). Repeated infections result in chronic inflammation of the tarsal conjunctiva of the upper eyelid. The chronic condition results in trichiasis, a shortening of the upper lid with in-turning of the eyelashes. The painful abrading of the cornea, if not corrected, results in corneal scarring, opacity, and blindness (Mabey *et al.*, 2003; Mecaskey *et al.*, 2003).

Human onchocerciasis, or river blindness, is a parasitic infection caused by the filarial worm, *Onchocerca volvulus*, transmitted by the female blackfly of the genus *Simulium*. The greatest mobidity is from skin disease, however, microfilariae can enter the eye where permanent visual damage can be caused after years of exposure (Burnham, 1998). Onchocerciasis causes an estimated 389,000 DALYs (WHO, 2008). The drug of choice for onchocerciasis control, ivermectin, is distributed by national control programmes under the auspices of the African Programme for Onchocerciasis Control (APOC) and the Onchocerciasis Elimination Program for the Americas. In APOC countries, treatment is administered via Community-Directed Treatment with Ivermectin programmes (CDTI) in areas where prevalence of palpable nodules exceeds 20%.

Table 1.2 Current treatment strategies for the major NTDs. MDA=Mass Drug Administration

Disease	Treatment used	Treatment delivery approach	Mass treatment prevalence threshold	Distribution via	Reference
Soil-transmitted helminthiasis	Albendazole/ mebendazole	Annual MDA to school aged children and high risk groups in entire districts	20%	School	(WHO, 2006c)
Schistosomiasis	Praziquantel	Annual MDA to school aged children and high risk groups to high risk schools	10%	. School	(WHO, 2006c)
Lymphatic filariasis	Single doses of albendazole plus ivermectin or single doses of albendazole plus DEC	Annual MDA for at least five years.	1%	Community	(WHO, 2005)
Trachoma	Surgery, Antibiotics (azithromycin), Facial cleanliness and Environmental change (SAFE).	Annual MDA for at least 3 years to entire districts	10%	Community	(WHO, 2006b)
Onchocerciasis	Ivermectin	Annual MDA	20%	Community	(Noma et al., 2002)

1.3 PARASITE LIFE CYCLES

The following section gives an overview of the life cycles of the parasites included in this thesis. An understanding of the fundamental biological features of parasites is central to understanding spatial patterns of diseases which, in turn, should inform the optimal spatial scale at which they are controlled.

1.3.1 Soil-transmitted helminths

The life cycles of STH, *S. mansoni* and *P. falciparum* are schematically depicted in Figure 1.1. In contrast to *S. mansoni* and *P. falciparum*, STH have direct life cycles such that they do not include intermediate hosts or vectors (Brooker and Bundy, 2008). STH life cycles vary slightly between species with eggs being the infective stages of *A. lumbricoides* and *T. trichiura* and free living larvae that hatch from eggs being the infective stages of hookworm. In addition, infection with *A. lumbricoides* and *T. trichiura* occurs through ingestion of eggs, whereas hookworm larvae mainly directly penetrate the skin. As STH life cycles do not rely on an intermediate host/vector, it could be expected that infection with STH species would be relatively widely distributed in space, as infection is only limited by the distribution of free living stages and human hosts.
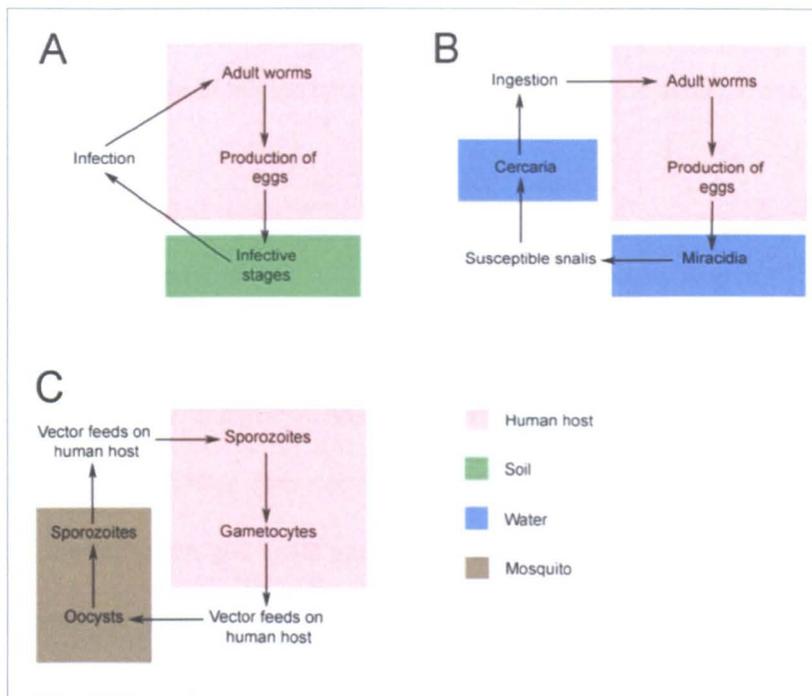
Figure 1.1 Simplified representation of the life cycles of (A) STH, (B) *S. mansoni*, (C) *P. falciparum* (adapted from Anderson & May 1991).

1.3.2 Schistosoma mansoni

Schistosoma spp life cycles differ from those of STH in two major ways: first, the requirement of an intermediate host, *Biomphalaria* snails in the case of *S. mansoni*; second, the inclusion of two free living stages, miracidia which are released from eggs in contact with water, and cercariae which are shed by infected snails (Figure 1.1). The life cycle is completed when a host is found, the cercariae penetrate the skin, migrate in the blood via the lungs to the portal vein, and transform into adults which mate and migrate to their perivesicular or mesenteric destination (Ross *et al.*, 2002; Gryseels *et al.*, 2006). The reliance on an intermediate freshwater snail host creates a

clustered spatial distribution of infection (Brooker, 2007), as transmission is only possible at water bodies that allow the overlap of human hosts, free living miracidia and cercaria, and the snail.

1.3.3 Plasmodium falciparum

The life cycle of *P. falciparum* is more complicated and involves a mosquito vector and no free living stages. Parasites enter the human host in the form of sporozoites, which undergo a number of liver and blood stages before gametocytes are formed which are able to reinfect mosquito vectors. Inside the mosquito gut, male and female gametocytes reproduce before forming an oocyst which develops over the following 1-3 weeks, eventually producing sporozoites which migrate to the salivary glands ready to infect another human host. Such a life cycle means that transmission is only possible where conditions suit the overlap of human hosts, mosquito vectors and parasite. In much the same way as *S. mansoni*, therefore, infection with *P. falciparum* is likely to cluster over smaller scales than STH as transmission is constrained to areas that suit the survival and development of host, parasite and vector.

1.4 SPATIAL ECOLOGY OF INFECTION

Infectious agents are not randomly distributed in space. Over large scales, parasites, as well as intermediate and definitive hosts, are constrained in their spatial distribution by the availability of suitable ecological and environmental niches. For example, prevalence of STH, *S. mansoni* and *P.*

falciparum typically displays a concave relationship with temperature showing slight differences in thermal limits (Craig *et al.*, 1999; Hay *et al.*, 2000; Brooker *et al.*, 2006b; Brooker, 2007). This relationship is presumably due to the effect of temperature on rates of parasite development as well as rates of survival, development and feeding interval of intermediate hosts (Sturrock, 1993; Craig *et al.*, 1999; Shope, 1999; Hay *et al.*, 2000; Tun-Lin *et al.*, 2000; Guerra *et al.*, 2008). Rainfall is also known to affect transmission of these parasites (Sturrock, 1993; Craig *et al.*, 1999; Brooker and Michael, 2000; Hay *et al.*, 2000; Kleinschmidt *et al.*, 2000), which is likely due to increasing availability of suitable habitats for free living infective stages, intermediate hosts and vectors in wetter areas, although excess rain can result in parasites being flushed from habitats or can cause the cessation of development in waterlogged soils (Brooker and Michael, 2000). Distance to water bodies has also been shown to be an important risk factor for *S. mansoni* infection due to overlap of human host and infective freshwater snails (Clements *et al.*, 2006a; Clements *et al.*, 2008b). Several other environmental factors such as soil-type, pH, water body type, humidity and vegetation density have all been linked to transmission of these parasites over varying scales (Thomson *et al.*, 1999; Kleinschmidt *et al.*, 2000; Mabaso *et al.*, 2003).

As a result of these ecological associations, diseases are spatially heterogeneous, albeit at varying scales. Understanding and quantifying the degree of spatial heterogeneity that diseases display is crucial to the design of surveys, as it influences the scale at which control and interventions should be carried out. Whilst a number of studies have looked at spatial heterogeneity at a micro- (village level) scale (Kloos *et al.*, 1998; Utzinger *et al.*, 2003; Brooker *et al.*, 2004a; Clennon *et al.*, 2004; Brooker *et al.*, 2006a), or have included the effect of clustering in statistical analyses of risk factors

(Clements *et al.*, 2006a; Kazembe *et al.*, 2006; Raso *et al.*, 2006a; Raso *et al.*, 2006b; Clements *et al.*, 2008b; Raso *et al.*, 2009), the following section focuses on work that has been specifically tailored to looking at spatial heterogeneity of disease at the meso- (country) scale, as this is most relevant to the design of national control programmes. The main analytical tool used to investigate spatial heterogeneity is the semi-variogram.

Using semi-variograms, which describe the mean difference in prevalence between pairs of points as a function of the distance between points, Brooker *et al.* (2004b) showed that in Uganda, hookworm typically displayed clusters of up to 123 kms, with *A. lumbricoides* clustering up to 33 kms and *T. trichuria* showing no evidence of clustering. The more widespread distribution of hookworm is presumably due to motile infective larval stages which are able to migrate below the surface of the soil to avoid desiccation, thereby enabling survival in warmer conditions (Beaver, 1953; Udonsi *et al.*, 1980). The scales over which STH infection clusters as a group (i.e. infection with any STH species) will therefore depend on the prevalence levels of each of these three parasites and the degree of overlap displayed. Data suggest that in East Africa, hookworm is more prevalent than either *A. lumbricoides* or *T. trichuria* (Brooker *et al.*, 2009b), whereas in West Africa hookworm appears to be the least prevalent STH (Ratard *et al.*, 1991; Ratard *et al.*, 1992; Brooker *et al.*, 2006b). Assuming similar spatial processes of STH species occur throughout SSA to those seen in Uganda, it might therefore be expected that STH would display more focal distributions in West Africa than East Africa. Given the apparent differences in spatial characteristics of the different STH species and varying levels of prevalence throughout SSA, spatial analyses of STH over a number of settings are clearly required.

Again using semi-variograms, Brooker (2007) showed that *S. mansoni* displayed clusters of around 50 kms across a range of settings. As mentioned above, this relative focality is most likely due to the reliance of an intermediate freshwater snail host, which is supported by associations between transmission and distance to water bodies (Clements *et al.*, 2006a; Clements *et al.*, 2008b). From a survey perspective, the more widespread distribution of STH implies that integrating STH surveys with those for schistosomiasis should be possible as any spatial sampling method developed for *S. mansoni* will sufficiently capture the larger scale spatial heterogeneity of STH infection.

Research tailored specifically to understand the scales over which malaria clusters has yet to be comprehensively carried out. Kazambe *et al.* (2006), found spatial autocorrelation occurring up to over 2 decimal degrees (~220 kms), however, no attempt was made to remove large scale trends as the purpose of the study was to identify risk factors. Likewise, whilst Hay *et al.* (2009) generated semi-variograms for *P. falciparum* to justify splitting prevalence data by continent before geostatistical analyses, no estimates of the scales over which infection clusters were made as these were not necessary for the purpose of the study. Whilst other studies have generated estimates of spatial autocorrelation parameters for *P. falciparum* (Kleinschmidt *et al.*, 2000; Gemperli, 2003; Gosoniu *et al.*, 2006), these are done within a modeling framework which accounts for a number of covariates such as rainfall, vegetation density and temperature. Accounting for these covariates removes spatial autocorrelation from the data and therefore cannot be used to comment on the inherent spatial heterogeneity of infection.

Whilst an understanding of the spatial variation of infection and disease is crucial to the design of cost effective surveys and control interventions, it is important to appreciate the effect of aspatial variation. As mentioned earlier, within a given region, prevalence of infection is likely to vary. Some of this variation can be partitioned as spatial variation, i.e. patches of higher or lower prevalence caused by factors that are themselves clustered. The remaining variation may be distributed randomly between sites and is therefore 'noise' or 'nugget' variation. Some of this noise may be attributed to sampling error, however, there are also likely to be other factors that influence the prevalence of infection which are themselves randomly distributed, such as differences in personal hygiene between sites. The presence of aspatial variation means that whilst infection may cluster over large scales, within the district infection levels could vary significantly due to non-spatial factors. In such settings, decision making at district level may not be appropriate. Equally, a disease may cluster over very small scales, but overall variation of infection is low (for example, prevalence could consistently be between 10-20%). In this case, making decisions on intervention at district level may be justified. As well as spatial and aspatial variation, prevalence may have important implications for the scales over which surveys are conducted. In high prevalence settings fine resolution mapping to identify foci of infection will not be required, irrespective of the scales over which infection clusters, as the majority of sites will likely qualify for intervention. Despite their importance, few studies have explored the implications that spatial and aspatial variation and prevalence have on the choice of survey, and intervention, scale.

1.5 EPIDEMIOLOGY

A consideration of the spatial patterns of infection and disease are required in the design of surveys. Equally important, from a survey perspective, is an understanding of the epidemiology of diseases, which inform the methods used to measure transmission. Parasite transmission is quantified on the basis of the basic reproductive rate R_0, which, for microparasites such as *Plasmodium* spp, is defined as the number of secondary infections caused by the introduction of a single infectious case into a completely susceptible population (Anderson and May, 1991; Smith *et al.*, 2007b). For macropasites, such as STH, R_0 is defined as the average number of offspring produced by a parasite over its reproductive lifespan that themselves survive to reproductive maturity in the absence of density dependent constraints on population growth (Anderson and May, 1991). As R_0 is both a measure of transmission and a measure of the effort required to eradicate transmission, it provides an ideal index for planning control programmes. Despite this, R_0 is rarely recorded due to logistical and computational difficulties in its estimation, particularly for malaria parasites (Smith *et al.*, 2010).

For STH and *Schistosoma* species, worm burden provides a direct measure of transmission since worms are the unit of transmission. In addition, studies have shown that worm burden is related to the degree of morbidity experienced by the human host (Arap Siongok *et al.*, 1976; Stephenson *et al.*, 2000). Direct estimation of mean worm burden is, however, relatively labour-intensive for STH and impossible for *Schistosoma* species. As such, quicker methods of assessing transmission have been investigated. Worm expulsion techniques have shown a quantifiable relationship between worm burden and egg output (Anderson and May, 1991), allowing the estimation of

worm burden possible via faecal egg counts. Furthermore, studies show that for STH and *S. mansoni* there is a predictable non-linear relationship between intensity and prevalence of infection (Guyatt *et al.*, 1990; Guyatt and Bundy, 1991) (Figure 1.2).
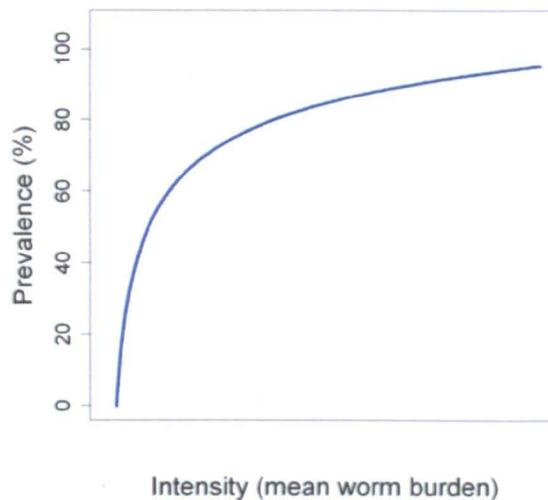
Figure 1.2 Typical relationship between prevalence and intensity of helminth infections (adapted from Guyatt *et al.* 1991). Note that the mean worm burden is not shown along the x axis due to differences between species.

This relationship shows that in areas of high prevalence, minor changes in prevalence can have considerable impacts on intensity of infection and hence morbidity. Furthermore, prevalence of infection, a much easier measurement than intensity, can be used as a proxy measurement of intensity, and therefore transmission, and is now the recommended survey measurement for a number of infectious diseases including STH and *S. mansoni* (WHO, 2000b, 2006c).

The gold standard measure of the intensity of malaria transmission is the Entomological Inoculation Rate (EIR), which is the product of the mosquito biting rate and the number of mosquitoes with sporozoites in their salivary glands. EIR is notoriously difficult to obtain, however, and is therefore rarely recorded. Conversely, prevalence of infection (also called parasite rate, PR) is widely used to determine infection risk and was a key measurement index during the Global Malaria Eradication Programme between 1950 and 1975 (Hay *et al.*, 2008). Smith *et al.* (2005) provide an approximation of the relationship between PR and EIR, allowing transmission intensity to be estimated from historic estimates of prevalence.

Both measurements of *Plasmodium* parasite prevalence and EIR suffer drawbacks, however. In particular, prevalence can fluctuate seasonally and is affected by exposure-related immunity, as well as drug use and resistance (Corran *et al.*, 2007). Similarly, measurements of EIR are affected by marked heterogeneity in the distribution of mosquitoes and very low sporozoite infection rates, even in highly endemic regions (Mbogo *et al.*, 1995; Drakeley *et al.*, 2003; Mbogo *et al.*, 2003). Serological methods, may, in part, be able to overcome these difficulties in order to provide a robust measurement of transmission as the antibodies they detect can persist for some time, allowing seasonal affects to be smoothed out (Corran *et al.*, 2007). Conversely, the fact that antibodies can persist long after infection, makes this approach inappropriate for use as a diagnostic test at the individual level. There may, however, be an opportunity to tailor serological tests to specific settings with the use of different serological markers which vary in their immunogenicity or target different age groups which would allow temporal changes to be elucidated (Drakeley *et al.*, 2005; Corran *et al.*, 2007). In addition to the fact that serological

methods are inappropriate for diagnosis at the individual level, the lack of standardized cutoff values for seropositivity and potential cross reactivity with antigens from other infectious agents (Abramo *et al.*, 1995), are challenges that need to be overcome. Due to their relative ease, particularly with the advent of rapid diagnostic tests, point estimates of parasite prevalence are still the preferred measurement of malaria transmission (Hay *et al.*, 2008).

The use of prevalence as a survey measurement requires an understanding of differences in risk between individuals within a community, as surveying those groups with the highest prevalence of infection provides information on those most at risk and is therefore the most sensitive method to detect transmission. It is now clear that prevalence displays marked age-dependent patterns across parasite species. For STH, maximum prevalence of *A. lumbricoides* and *T. trichiura* is usually attained by the age of five, whereas hookworm reaches a peak in adolescence or early adulthood (Hotez *et al.*, 2006a) (Figure 1.3). For *S. mansoni*, infection prevalence typically increases from an early age, peaking in teenage years and decreasing throughout adulthood (Jordan and Webbe, 1993; Kabatereine *et al.*, 2004). Whilst dependent on transmission setting, malaria prevalence tends to peak in early childhood, decreasing with increasing age (Smith et al., 2007a; Brooker et al., 2009c) (Figure 1.3).

These epidemiological differences between parasite species have important consequences for the design of integrated surveys. The age-prevalence profiles mean that targeting primary school children (typically aged between 5-15 years), not only offers a practical unit in which to conduct

surveys for these diseases, but provides an epidemiologically sensible age group to target. This age group has historically been targeted for STH and *S. mansoni* surveys, and more recently for malaria (Brooker *et al.*, 2009c; Gitonga *et al.*, 2010).
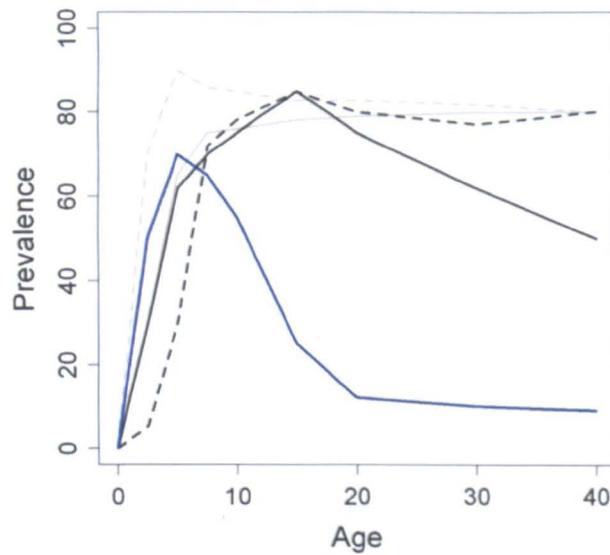
Figure 1.3 Typical age-prevalence curves for hookworm (black dashed), *T. trichiura* (grey dashed), *A. lumbricoides* (grey), *S. mansoni* (black) and *P. falciparum* (blue) (adapted from Hotez *et al.* (2006a) and Smith *et al.* (2007a)).

1.6 DIAGNOSTIC TECHNIQUES

In order to produce accurate information on the spatial distribution of disease, surveys are heavily dependent on the reliability of the diagnostic methods used. Diagnostics are rarely perfect and, as such, the choice of diagnostic test has to be based on its performance, cost and ease of use. The

following section summarises the main diagnostic tests used in large scale surveys for the parasite species in question.

1.6.1 STH diagnostics

Due to the relationship between prevalence and transmission, as discussed above, the standard diagnostic approach for STH and *S. mansoni* is the detection of eggs in faeces. Based on ease and relative low cost, the Kato-Katz technique, which uses duplicate 41.7 mg faecal smears stained with malachite green, is the most commonly used diagnostic method (WHO, 1991). Due to concerns over sensitivity of this method in low intensity infections (Booth *et al.*, 2003), other diagnostic methods have recently been (re)explored for STH species. Several studies have shown that using the FLOTAC apparatus, which allows separation of the floating suspension carrying eggs following centrifuging, yields a higher sensitivity than Kato-Katz, even if triplicate Kato-Katz smears are examined, but results in lower egg concentrations (Utzinger *et al.*, 2008; Knopp *et al.*, 2009; Glinz *et al.*, 2010). That said, by pooling results from all diagnostic methods these studies demonstrate that FLOTAC is not able to detect all positive stool samples, highlighting the potential benefits of using a combination of diagnostic techniques in low transmission settings, where resources allow (Knopp *et al.*, 2008). Despite these results, Kato-Katz is still recommended by WHO for STH diagnosis because of the high costs of FLOTAC.

1.6.2 Schistosoma spp. diagnostics

Whilst stool examinations have been the cornerstone of *S. mansoni* diagnosis, several Rapid Diagnostic Techniques (RDTs) have been investigated (Stothard, 2009). The rapid assessment of

urinary schistosomiasis has been aided by the development of school-based questionnaires that establish history of haematuria (blood in urine), a diagnostic feature of the infection (Lengeler *et al.*, 1991; Lengeler *et al.*, 2002). The similar use of questionnaires for diagnosis of intestinal schistosomiasis was thought to be possible given previous findings of an association between *S. mansoni* infection and blood in stool and/or bloody diarrhoea (Ongom and Bradley, 1972; Lengeler *et al.*, 2002). Unfortunately, results collated from the field indicate only moderate sensitivity, and as such, validation studies are recommended in any given setting before use (Lengeler *et al.*, 2002).

More recent diagnostic developments include antigen capture dipsticks which detect schistosome circulating cathodic antigen (CCA), a mixture of proteins released from the gut of adult worms, in urine samples. Studies indicate a high sensitivity and specificity to *S. mansoni* infection over a number of settings (Stothard *et al.*, 2006; Legesse and Erko, 2007; Standley *et al.*, 2010), although difficulties in interpreting borderline results mean that further modifications may be required (Standley *et al.*, 2010). Investigations into the performance of CCA dipsticks for the detection of *S. haematobium* infection have generated poor results ranging from a complete failure to detect infection (Stothard *et al.*, 2006), to low sensitivity and specificity of 52% and 62% respectively (Ayele *et al.*, 2008). Recent improvements to the technique appear, however, to have produced more acceptable levels of performance (88 – 96% sensitivity) (Midzi *et al.*, 2009). Unfortunately, CCA antigens are genus cross-specific making differentiation between the two species impossible - a challenge that needs to be overcome before its use in the field (Stothard, 2009). Additionally, these tests are currently relatively expensive, retailing at between $2.60 and $4.60 and, as such,

Kato-Katz is still the recommended diagnostic approach (WHO, 2006c). Another method, the soluble egg antigen enzyme-linked immunosorbent assay, has also shown good levels of sensitivity (89%) and specificity (70%) to *S. haematobium* infection and holds promise as a field based detection method (Stothard *et al.*, 2009).

1.6.3 P. falciparum diagnostics

Traditionally, the gold standard for malaria diagnosis has been microscopic examination of Giemsa-stained blood smears (Hay *et al.*, 2008). Due to the need for highly skilled technicians, several other methods of identifying infection have been developed. RDTs which detect malaria antigens in very small volumes of blood, have become popular as a method of parasite detection due to their high sensitivity and relatively low cost (Murray *et al.*, 2008) and there are now an estimated 60 different brands available commercially (WHO, 2009b). A WHO report on malaria RDT performance concluded that whilst a small number of tests demonstrated consistent detection of parasites over a range of concentrations, are stable in tropical temperatures and are easy to use, many failed to detect parasites, particularly when at low blood concentrations (WHO, 2009a, b). Additionally, the report found variation between lots and similar products, highlighting the need for lot checking post purchase and in the field before use. Recent advancements in RDT technology have also allowed *P. falciparum* infections to be distinguished from other *Plasmodium* infections. Other methods, such as Polymerase Chain Reaction, have been used to detect presence of infection and, although more costly and complex, may prove invaluable in low transmission settings and during the evaluation of control programmes where density of parasites may be low and missed by conventional methods (Okell *et al.*, 2009).

1.7 SURVEY DESIGN

The previous section has given an overview of the current diagnostic methods available to identify infected individuals. In most cases, however, it is not possible to test every individual and population-based surveys are required to gain an understanding of the distribution of infection and disease and to allow targeted use of interventions. For large scale surveys of human populations, simple random sampling, i.e. a random selection of individuals from the entire population, is rarely used due to the difficulties in identification of all individuals required to derive a complete sampling frame, and the financial and logistical costs associated with reaching all selected individuals. These constraints have led to the use of survey designs where sampling frames are constructed that identify groups or clusters of enumeration units. Such cluster sampling minimizes operational and financial costs and is generally considered to be the gold standard survey method for providing prevalence estimates within regions.

Cluster survey methods are recommended by WHO for STH and *S. mansoni* surveys (Montresor *et al.*, 1998). First, the country or district is divided into ecologically homogeneous areas. Within those areas where transmission is suspected, five to ten schools should be chosen. In each school, fifty children are selected from any of the three upper classes and stool samples examined using Kato-Katz method (WHO, 2006c). This survey technique is based on previous suggestions that this sample size should provide adequate precision of prevalence and intensity (Lwanga and Lemeshow, 1991; Montresor *et al.*, 1998). These methods, however, may not be entirely suitable for disease control programmes for a number of reasons. Firstly, it is difficult to define the ecological zones required for stratification. Secondly, interventions are often carried out within

administrative units such as districts or sub-districts, and translating results based on ecological zones may be problematic. Thirdly, whilst precise estimates of prevalence are useful for the evaluation of an intervention, from a decision making perspective, a simple understanding of whether prevalence exceeds a given threshold would suffice. For example, Mass Drug Administration (MDA) of praziquantel is recommended in areas where prevalence of schistosomiasis is ≥10%. When deciding whether an intervention should go ahead or not, an estimate of whether prevalence is ≥10% is therefore as useful as a precise estimate of prevalence, but will most likely require less sampling effort.

Surveys for malaria typically use designs which assess the prevalence of parasitaemia in young children through household cluster surveys as part of Malaria Indicator Surveys (MIS) (Roll Back Malaria Monitoring and Evaluation Reference Group, 2005) or Demographic and Health Surveys (www.measuredhs.com). Survey recommendations are flexible allowing incorporation of strata based on ecological zone or urban/rural status and varying sample sizes based on economic and logistical considerations. If recorded, information on malaria infection and anaemia is typically taken from children under the age of 5 years. This age group is, however, suboptimal as infection prevalence in very young children is modified by a number of factors such as maternal antibodies (Hviid and Staalsoe, 2004). Furthermore, these types of national cluster surveys are expensive to conduct, time-consuming and are powered only to provide country/province level prevalence estimates. These drawbacks make this method unsuitable for regular monitoring and evaluation and don't allow decisions on control to be made at smaller administrative levels such as district or sub-district.

School based surveys have also been used to describe the distribution of prevalence at national and regional level, and may provide cheaper and more efficient methods to conduct surveys and surveilance (Brooker *et al.*, 2009c; Gitonga *et al.*, 2010). These survey designs, however, in a similar way to those for STH, are aimed at providing prevalence estimates. In terms of guiding intervention, it may be possible to simplify surveys in order to classify populations according to prevalence thresholds used to decide intervention strategy.

1.7.1 Rapid assessment

Rapid assessment techniques are a group of survey methods which can be differentiated from population based surveys by the fact that they aim to provide a rapid method to classify areas according to intervention need. Such techniques now exist for a number of health problems (Anker, 1991; Vlassoff and Tanner, 1992; Brooker *et al.*, 2009a) (Table 1.3). For example, the use of simple school-based questionnaires has been shown to be an inexpensive valid method of identifying communities at high risk of urinary schistosomiasis (Lengeler *et al.*, 1991; Lengeler *et al.*, 2002; Clements *et al.*, 2008a). This method uses teachers to administer questionnaires to school children to assess whether they have recently experienced any of the symptoms associated with urinary schistosomiasis, particularly blood in urine. Results are then collated centrally to assess which schools qualify for mass treatment. This method has now been validated in a number of countries, however, local validation is advised before large scale use (Lengeler *et al.*, 2002).

Another approach for simplifying sampling that has received attention in a number of public health settings is Lot Quality Assurance Sampling (LQAS) (Lemeshow and Taber, 1991; Robertson and Valadez, 2006). Taken from techniques developed for the manufacturing industry in the 1920s, this method allows the categorization of populations based on prevalence of disease, using small sample sizes. The principle is that a small representative sample of a population is screened and if a certain number of individuals are found to be infected, the population is classified as high prevalence. The number of allowable infected individuals is based on pre-defined error rates and a statistically determined sample size. As the technique classifies populations into those that fall above or below certain threshold values, it is best used when making decisions on whether to intervene rather than as a means of calculating prevalence or intensity of infection (Brooker *et al.*, 2005). LQAS has now been employed in a number of health related situations including the monitoring of vaccination coverage (Lanata *et al.*, 1990; Singh *et al.*, 1996) and leprosy elimination monitoring (Gupte *et al.*, 2004), as well as helping to guide control of *Trypanosoma brucei* (Hutin *et al.*, 2004), and trachoma (Myatt *et al.*, 2003; Myatt *et al.*, 2005) (Table 1.3).

Brooker *et al.* (2005) showed that LQAS provides a simple rapid assessment technique for *S. mansoni*. Computer simulations showed that with 15 children per school it was possible to classify >90% of schools correctly, according to WHO prevalence thresholds, with field studies showing more variable but similarly encouraging results. By incorporating estimates of survey and treatment costs, the authors were also able to show that employing LQAS was more cost-effective than mass treating all schools in settings where prevalence of disease was <75%. Whilst LQAS reduces sampling effort at each school, it is likely to be difficult to carry out over large scales and

further work into survey designs that provide high resolution data with minimal sampling effort are still required.

For malaria, LQAS has been used in Mozambique to provide regional information on bednet distribution using data collected during national MIS (Biedron et al., 2010). In terms of disease prevalence estimates, LQAS has also been used in Madagascar to confirm routine surveillance reports, collected by primary health centres, of increased malaria transmission (Rabarijaona et al., 2001). This study showed that when trying to identify schools with a prevalence of ≥15%, sampling 36 children from a school with a stopping rule of two (i.e. if two or more children were found positive the school was classed as high prevalence), provided a sensitivity of 100% and specificity of 86%, when compared to a gold standard sample of 70 children. Whilst this study shows that the use of LQAS to confirm individual reports of transmission is operationally feasible, it does not assess whether LQAS can be used on a large scale. Furthermore, the prevalence threshold of 15% used in this study differs from the recent recommended threshold of 5% (Hay et al., 2008), a proposed prevalence level at which to switch to elimination oriented surveillance and control, and no attempt was made to incorporate cost into the analyses. An exploration of different survey types, using the 5% threshold and including cost estimates, is clearly required.

Table 1.3 Summary of existing rapid assessment methods for the major NTDs and malaria.

Disease	Sampling method	Age group typically surveyed (years)	Sample size	Implemen-tation unit	Method of diagnosis	Threshold prevalence for intervention
Soil-transmitted helminthiasis	N/A	9-15	N/A	District	Microscopic examination of stool samples	20%
Urinary schistosomiasis	School-based questionnaire (Lengeler et al., 2002)	9-15	Variable[1]	School	Questionnaire	10%
Intestinal schistosomiasis	School-based LQAS (Brooker et al., 2005)	9-15	15	School	Microscopic examination of stool samples	10%
Lymphatic filariasis	RAGFIL (Gyapong and Remme, 2001; Gyapong et al., 2002)	>15	50-100	District	Filarial antigen surveys	1%
Trachoma	Trachoma Rapid Assessment (Negrel et al., 2001; WHO, 2006b)/ LQAS (Myatt et al., 2005; Faye et al., 2006)	1-9 (TRA) 2-5 (LQAS)	50	District	WHO simplified trachoma grading scheme	10%
Onchocerciasis	REMO (Noma et al., 2002)	>20 (males preferred)	50	Varies	REA, based on nodule presence	20%
Loaisis	RAPLOA (Takougang et al., 2002; TDR, 2002)	>15	80	Community	Questionnaire	No MDA of ivermectin if >40%
Malaria	LQAS (Rabarijaona et al., 2001)	5-12	36	School	Microscopic examination of blood slides	15%[2]

[1] All children in 1 class from each of three grades covering the age range 9-15.
[2] Arbitrary threshold used for the purpose of this study
LQAS – Lot Quality Assurance Sampling, **RAGFIL** – Rapid Geographical Assessment of Bancroftian Filariasis, **REMO** – Rapid Epidemiological Mapping of Onchocerciasis, **RAPLOA** – Rapid Assessment Procedure for Loa loa, **REA** – Rapid Epidemiological Assessment, **MDA** – Mass Drug Administration.

LQAS traditionally relies on simple random sampling for data collection. Recently, however, LQAS in combination with cluster sampling has been tested to classify entire areas based on prevalence of malnutrition (Deitchler *et al.*, 2007; Deitchler *et al.*, 2008). These studies showed that a 33 x 6 (33 clusters, 6 individuals per cluster) and 67 x 3 LQAS design can be used to determine whether prevalence of acute malnutrition exceeds a critical threshold. When compared to a conventional 30 x 30 design results showed that LQAS surveys were substantially cheaper and quicker to carry out, offering a useful alternative in emergency settings. Such methods provide opportunities for directing interventions cost-effectively in a number of settings, however, without accounting for between cluster correlation, error rates of such methods can be affected (Hedt *et al.*, 2008; Olives *et al.*, 2009). From an integrated survey perspective, this type of survey technique is not an obvious candidate, owing to the use of stopping rules whereby surveys continue until a threshold target is met. Having a number of different stopping rules for the different diseases might make an integrated approach more operationally complex.

Other NTD rapid assessment techniques include the Rapid Geographical Assessment of Bancroftian Filariasis (RAGFIL). As a first step, existing data and information are collated to help classify implementation units (IUs) as those with presence of transmission, no transmission or highly unlikely and transmission likely but unknown. In areas where transmission is likely but unknown, surveys are carried out on adults from communities selected by a 50 km x 50 km grid using rapid diagnostic Immunochromatographic Card Tests (ICT) which detect circulating antigens (WHO, 2005). Results are then fed back into GIS packages and together with spatial statistics, contour maps of prevalence are generated. These data are then used to classify the IU as an area

with transmission >1%, or no/unlikely transmission. The RAGFIL method has been used successfully to estimate the distribution of LF in four countries in West Africa (Gyapong *et al.*, 2002). Such a spatial survey methodology, in combination with spatial prediction, may offer an opportunity to reduce the number of survey sites required and could prove useful for other parasite surveys. The spatial analysis on which RAGFIL is based, however, does not take into account large scale trends which may affect conclusions on the scales over which LF clusters (WHO, 1998; Gyapong and Remme, 2001). Indeed, recent studies, albeit carried out outside SSA, have raised concerns that small foci of infection may persist between interstices of a 50 × 50 km grid (Srividya *et al.*, 2002; Alexander *et al.*, 2003). A re-examination of the scales over which LF clusters is required in order to inform the optimal spatial resolution of LF mapping and control.

APOC uses the Rapid Epidemiological Mapping of Onchocerciasis (REMO) procedure to direct control towards those most at risk (Noma *et al.*, 2002). As a first stage, the country is divided into different biogeographical zones, based on the density, spatial distribution and host seeking behaviour of the blackfly vector. As these vectors seldom fly more than 15km in search of a blood meal, the majority of transmission occurs close to their preferred breeding sites of fast flowing and well oxygenated rivers and streams (De Sole *et al.*, 1991a; De Sole *et al.*, 1991b). Based on this, two types of community are selected in each zone: 'high risk' and 'secondary' (located at least 10km from the probable main source of vectors). Each high risk community is then surveyed using the rapid-epidemiological-assessment, based on nodule presence in a random sample of 50 adult males. If high-risk communities are found to be meso- or hyper-endemic for onchocerciasis, the secondary community is surveyed to provide a better understanding of the distribution and

severity of the disease. This information is then collated using GIS to visualise priority areas for ivermectin distribution. Definite-CDTI communities are those where prevalence of nodules is >20%. Communities where prevalence is <20% are classified as no-CDTI and clinic-based ivermectin treatment may be provided. Possible-CDTI areas are those where epidemiological information is not clear enough for classification and further surveys may be required. REMO has now been completed in most APOC countries and control programmes exist in nearly all known meso- and hyper-endemic areas (Noma et al., 2002; Etya'ale, 2008). Further large scale mapping for directing onchocerciasis control is therefore no longer a priority, however, surveillance measuring the impact of interventions and need for further control is still required.

Despite developments in rapid assessment techniques for a number of NTDs, no such framework exists for the assessment of STH. This is most likely due to the fact that STH are considered to be relatively spatially homogeneous and drug treatment is thought to be cheap enough to warrant mass treatment without the need for surveys. In terms of the development of a rapid assessment method for these parasites, there are a number of issues that need to be addressed (discussed in chapter 2). Indeed, more generally, the influence of spatial heterogeneity and cost (both in terms of survey and treatment cost) on the design of disease surveys is not well understood.

1.7.2 Geostatistical survey design

As emphasized in previous sections, there is a need to investigate the influence of spatial heterogeneity on the design of disease surveys. It may also be possible to go beyond simply accounting for spatial characteristics of diseases, to developing spatially informed survey designs that, in a similar way to RAGFIL, allow infection prevalence to be predicted at unsurveyed sites using geostatistical techniques. Such geostatistical survey designs are widely used in natural resource, agricultural and fisheries surveys (Van Groenigen et al., 1998; Stein and Ettema, 2003), as they help to reduce the sampling effort required whilst enabling high resolution information. Other than in the RAGFIL design, the application of geostatistical survey methods to the design of parasitological surveys is surprisingly lacking.

Geostatistics is centered around the general principle that 'everything is related to everything else, but near things are more related than distant things' (Tobler, 1970). In statistics, this phenomenon is known as positive spatial autocorrelation, and is a common observation of variables across all scales of geographic space (spatial autocorrelation), as well as time series (temporal autocorrelation). Despite the apparent self-evidence of this observation, only recently have the implications of spatial autocorrelation on ecological studies been investigated (Legendre, 1993; Thomson et al., 1999; Legendre et al., 2002). Although there are important repercussions for the design and analyses of studies investigating associations between spatially autocorrelated data due to the non independence of data (see Legendre et al. (2002)), this section will focus on the implications that geostatistics has for survey design.

Infection and disease display spatial autocorrelation due to the clustered nature of appropriate ecological niches suitable for parasite transmission (as discussed in section 1.4). In reality, all infectious agents are likely to be focal to some degree, a fact which is often overlooked in the literature, with more widespread diseases simply clustering over larger scales. Scale is an important consideration when understanding spatial autocorrelation. Values are related in space as a result of large (macro-), medium (meso-) and small (micro-) scale variations in factors that affect transmission. Macro-scale variation manifests as trends across large geographic areas, such as a gradient across a continent. This could, for example, be due to broad differences in temperature which affect the development and survival of pathogens (Gething *et al.*, 2011). Meso-scale variation describes the local spatial autocorrelation, also referred to as spatial heterogeneity. Such heterogeneity could be caused by local variations in suitable habitats for intermediate hosts or free living parasite stages. Micro-scale variation is brought about by differences within these meso-scale clusters such as the proximity of households to mosquito breeding sites (Bejon *et al.*, 2010).

From a survey design perspective, it is important to understand the degree of spatial autocorrelation at a scale of operational relevance. Macro-scale variation in risk can be generally predicted using climatic variables, such as temperature and rainfall, which can be used to target subsequent surveys. At the other end of the scale, whilst control interventions may be most cost-effective when targeted at the household level (Carter *et al.*, 2000), the operational requirements of mapping at this scale are likely to be unfeasible. National disease control programme managers are therefore most interested in the meso-scale spatial variation within a country. At this scale, all

things being equal, diseases that cluster over larger scales require fewer survey points than more focal diseases that require higher resolution data to avoid missing foci of infection. There is, therefore, a need to understand the distances over which diseases typically cluster at this scale.

In addition to establishing the scales over which parasites cluster, an understanding of spatial structure allows the use of spatial interpolation techniques such as kriging, which make use of spatial autocorrelation to predict values at unsurveyed locations based on values at surrounding sites (Oliver *et al.*, 1992; Kleinschmidt *et al.*, 2000; Pfeiffer *et al.*, 2008). A large body of work from the natural resources, agricultural and environmental sectors have explored optimal spatial survey designs that minimize kriging prediction error (kriging variance) (Lloyd and Atkinson, 1998; Van Groenigen and Stein, 1999; Lark, 2000; Van Groenigen, 2000; Stein and Ettema, 2003). In the presence of spatial autocorrelation, surveying sites that are situated close to each other is less useful than surveying sites further apart, as locations close to each other can be easily predicted. As sites get further apart, however, kriging variance at locations between survey points increases. To minimize kriging variance for minimal sampling effort, therefore, sites should be located at regular locations to avoid redundancy, but should be at high enough resolution to enable efficient spatial interpolation. Survey designs optimized for minimizing kriging variance, which would allow a reduction in the number of survey sites required, warrant further investigation in a disease context, particularly in resource poor settings where maximizing cost-effectiveness is paramount.

1.7.3 Survey cost considerations

A key feature of any survey design is defining the required level of precision. For surveys aimed at guiding interventions, deciding the required level of precision is not always straightforward and is often determined by the availability of resources, i.e. achieving the best level of precision given a set financial amount. Despite the popularity of clustered survey designs for epidemiological surveys, surprisingly few studies have investigated the cost implications of different sample sizes or survey strategies. Connelly (2003) provides an excellent example of optimizing cluster sample designs for randomised control trials. He describes the use of isoquants which show the different combinations of cluster number and size required to achieve a given level of precision. By estimating costs per cluster and per individual, it is then possible to identify the cheapest cluster design for a given level of statistical power. Similarly, Williams *et al.* (2008) describes a similar method to identify the most cost-effective cluster survey design for estimating prevalence of tuberculosis in Cambodia. By calculating the cost of different sized cluster survey designs, the authors provide an equation to calculate the cheapest cluster design that achieves a given level of precision.

In terms of surveys designed for intervention decision making, these studies have a major drawback, namely that costs of subsequent treatment – either appropriate or inappropriate - are ignored. Including this information is important as financial resources allocated for disease control are often used for both surveys and treatment. Under such a situation, where surveys are cheap relative to treatment, higher levels of precision should be attempted as the cost of misclassification (in terms of unwarranted administration of drugs) is very high. Where treatment

is relatively cheap, then from a financial perspective, less effort should be put into surveys and treatment can be administered more liberally. Equally, treatment safety should influence survey precision as drugs with negligible side-effects can be administered more liberally than more toxic drugs which should be targeted only at those infected individuals. Very few studies have, however, explored the relationship between survey and treatment cost, as well as treatment safety, and survey effort. Brooker *et al.* (2005) included the cost of treatment to compare a school by school screening and treatment with the cost of presumptive treatment in all schools without surveys. Findings suggested that targeting treatment using school surveys was more cost-effective where overall prevalence was <75%, whereas presumptive treatment was more cost-effective where prevalence was ≥75%. Similarly, in a study in Nigeria, Gutman *et al.* (2009) showed that presumptive treatment of school children was cheaper than a school by school screening and treating approach. This study, however, assumed screening involved sampling 30 children with several repeat trips to each school to distribute and collect stool samples as well as inform pupils of their infection status. As Brooker *et al.* (2005) show, there may be opportunity to screen schools for *S. mansoni* using sample sizes of 15 and examine slides in the field (discussed further in section 1.7.1). Such a method means that only one trip is required to each school, which considerably lowers survey costs.

These studies are important as they illustrate the cost-effectiveness of spatially targeted control of schistosomiasis, however, there is an opportunity to extend this analysis to investigate the cost and performance implications of different sampling methods (such as altering sample sizes) and classification techniques (such as choosing to overclassify borderline areas). This will allow an

investigation into optimal survey designs in settings where cost-effectiveness is more important than achieving set levels of precision. The design of rapid assessment surveys provides such an example.

1.7.4 Challenges and opportunities for integration

Despite the existence of a number of rapid assessment techniques, and the obvious interest in an integrated approach to surveys, few standard integrated disease survey protocols currently exist. Parasitic diseases differ in their biology, epidemiology, diagnostic techniques and recommended survey methods, which all complicate the simple combining of individual survey methodologies. Furthermore, surveys were developed at different times and for different purposes, without the intention of future integration. In addition, surprisingly little is known about the spatial heterogeneity of these diseases and the influence of cost on survey design. Diseases that are more focal, should require more survey effort than more widespread diseases. Equally, those diseases that are more expensive to treat, should require more survey effort than those for which there are cheap interventions, so as to minimize wasted resources in the form of unwarranted treatment. Despite the importance of these factors for survey design, their relative importance is not well understood. Thus, there remains a need for further theoretical and operational research on integrated surveys accounting for both spatial heterogeneity and cost.

The integration of surveys for STH and *S. mansoni* is aided by the fact that the recommended diagnostic technique for both parasites is examination of stool samples using Kato-Katz. Equally, the age-specificity of infection for these parasites is such that school-aged children are the

recommended survey group. The main challenge facing the integration of surveys for these parasites is the difference in spatial heterogeneity and IU. It is widely thought that surveys for *S. mansoni* are best done at school-level whilst surveys for STH can be done at district level. It could, therefore, be possible to conduct integrated surveys at school level as this would provide sufficiently high resolution information for both parasites. It is, however, unlikely that such an approach is operationally and financially feasible over large scales due to the large number of survey sites required. It would therefore be interesting to explore whether survey designs could be developed for *S. mansoni* that still allow decisions on control to be made at school level, using only a subset of schools. Using geostatistcs to inform the locations of survey sites that allow spatial prediction of prevalence at unsurveyed sites, may provide an option and warrants further investigation. Additionally, exploring the implications of carrying out surveys, and subsequent control activities, for *S. mansoni* over different spatial scales may help to evaluate whether there is a scale at which surveys can be integrated with those for STH and other diseases.

The integration of surveys for malaria with those for STH and *S. mansoni* is again aided by the fact that schools can be used as sample units (Brooker *et al.*, 2009c; Gitonga *et al.*, 2010; Ashton *et al.*, 2011). Differences to consider include contrasting spatial heterogeneity and the need for aggregate estimates of endemicity class over administrative units such as districts or sub-districts. An additional layer of complexity in optimizing malaria surveys is the incorporation of cost estimates. For STH and *S. mansoni*, estimating the cost of treatment (and mistreatment) is relatively straightforward, as there are clear guidelines on drug intervention with good estimates

of drug and delivery cost. In contrast, malaria control is often multi-faceted and varies between settings, which complicates estimates of cost-effectiveness.

1.8 AIMS AND OBJECTIVES

The overall aim of this thesis is to investigate and quantify the spatial heterogeneity of STH, *Schistosoma mansoni* and *Plasmodium falciparum* infection in East Africa using spatial statistics and to identify optimal survey designs for guiding control on the basis of both observed spatial heterogeneities and survey costs. Satisfying this aim will have the practical consequence of providing cost-effective rapid survey methods to better target parasite control.

The specific objectives include:

- To quantify the spatial heterogeneity of STH, *S. mansoni* and *P. falciparum* in a range of epidemiological settings in East Africa

- To establish the most cost-effective rapid survey design for STH to classify districts for mass treatment based on cumulative prevalence, accounting for spatial heterogeneity as well as survey and treatment costs

- To investigate the use of alternative survey designs for *S. mansoni* that aim to classify schools to target praziquantel treatment

- To investigate the cost-effectiveness of carrying out STH, *S. mansoni* and *P. falciparum* surveys at different administrative levels to identify whether there is a spatial scale at which surveys for these parasites can be operationally integrated.

1.9 THESIS OUTLINE

The remainder of this thesis is organized as follows. Chapter 2 reports on recent field attempts to design and implement integrated surveys in Southern Sudan and Kenya, using existing survey recommendations for NTDs and malaria. In addition to describing the survey methods used, this chapter also highlights practical lessons learnt from the different approaches, and aims to identify scientific and operational research questions relating to the design of future integrated surveys.

Recognising the lack of work on STH survey design, chapter 3 evaluates the cost-effectiveness of different survey approaches for STH that aim to classify districts according to WHO recommended prevalence thresholds. This chapter makes use of geostatistics to quantify the spatial heterogeneity of STH and to generate fully enumerated, pseudo gold standard prevalence data at schools in Kenya, which are used in computerized simulations to test the performance and cost-effectiveness of different survey approaches.

Extending this methodology, chapter 4 investigates different survey methods for *S. mansoni*, which aim to identify schools that qualify for treatment, according to WHO recommended prevalence thresholds. Specifically, the chapter compares the performance and cost-effectiveness of LQAS, the current recommended rapid survey method, with a geostatistical survey approach which allows spatial prediction of endemicity class at unsurveyed locations based on a subset of schools.

Chapter 5 assesses the performance and cost-effectiveness of conducting surveys for STH, *S. mansoni* and *P. falciparum* in four provinces in Kenya, at three different spatial scales: school level; sub-district; and district. In addition to investigating the effect of scale, this chapter also assesses the effect of prevalence on the choice of survey approach, and attempts to establish whether there is a scale at which surveys for these three parasites can be integrated.

The final chapter briefly summarizes the major findings of this thesis, and discusses future challenges and opportunities for the development and implementation of integrated survey designs for this group of parasites.

1.9.1 Peer-reviewed publications

Chapters 2 and 3 have been published in modified versions in peer-reviewed publications (Sturrock *et al.*, 2009; Sturrock *et al.*, 2010) (see Appendix) and a slightly modified version of Chapter 4 is in press in International Health. Data arising from surveys conducted in Southern Sudan (reported in Chapter 2) have also been used in an additional publication (Robinson *et al.*, 2009).

CHAPTER 2 | INTEGRATED MAPPING OF NEGLECTED TROPICAL DISEASES: EXPERIENCE FROM SOUTHERN SUDAN AND KENYA

2.1 INTRODUCTION

Efforts to control NTDs need to be based on an empirical understanding of the geographical distribution and overlap of different NTDs. As highlighted in chapter 1, there are currently no published guidelines for integrated rapid assessment surveys for NTDs, despite a number of recent attempts in the field (Hopkins *et al.*, 2002; Emerson *et al.*, 2008; King *et al.*, 2009). As guidelines are currently being developed by the WHO, it is important to learn from initial attempts to conduct integrated mapping of NTDs. This experience will also help identify the relevant research questions which need addressing in order to develop an optimal integrated mapping strategy.

In terms of the current recommended survey designs, WHO currently recommends that the MDA need for LF elimination is determined through LQAS of up to 250 individuals in each intervention unit (typically a district or equivalent administrative unit) (WHO, 2000b). LQAS surveys should be preceded by a review of existing information on LF and rapid assessments through questionnaires, seeking information on the prevalence of clinical manifestations of *W. bancrofti* infection (hydrocele and lymphoedema) from key informants (WHO, 2000b). Based on questionnaire data it

should then be possible to demarcate areas as endemic, non-endemic or still undetermined. Unfortunately, however, some studies have shown that presence of clinical signs of disease is not necessarily indicative of active transmission (Michael *et al.*, 1994; Eigege *et al.*, 2003). For schistosomiasis and STH control, WHO recommends that 200-250 school-aged children are sampled in each ecological zone (Montresor *et al.*, 1998), although it is unclear how best to define these zones. Surveys for malaria typically assess the prevalence of parasitaemia in young children through household cluster surveys as part of MIS, Demographic and Health Surveys (www.measuredhs.com), or via school based surveys (Brooker *et al.*, 2009c). Results are then usually summarized over administrative areas such as districts.

To assess how currently available survey protocols may be practically combined in the field, this chapter reports on two initial attempts to design and carry out integrated surveys in Southern Sudan and Kenya, two countries who have recently begun integrated control of NTDs. Integrated control typically targets those diseases for which safe and effective Preventive Chemotherapy (PCT) is readily available. The main NTDs endemic in Southern Sudan are onchocerciasis, LF caused by *Wuchereria bancrofti* infection, soil-transmitted helminthiasis (caused by STHs: hookworms, *Ascaris lumbricoides* and *Trichuris trichiura*), schistosomiasis (due to *Schistosoma haematobium* and *S. mansoni*), and trachoma caused by *Chlamydia trachomatis*. The principal focus of NTD control is community-based treatment campaigns; very few schools currently exist in Southern Sudan precluding school-based approaches. In Kenya, the main NTDs are LF, STH, schistosomiasis and trachoma, with school-based delivery of treatment a major focus of

government efforts. In the Kenya surveys, malaria was additionally assessed because of the potential of integrating malaria control in current school health programmes (Brooker, 2009).

The primary aim of this chapter is to describe and evaluate the design and implementation of integrated NTD surveys in two contrasting settings in Africa and discuss the research gaps that require exploring to develop optimal integrated survey designs. Additionally, the implications of the survey findings for the design and delivery of integrated NTD control programmes are considered.

2.2 METHODS

2.2.1 Personal involvement

Surveys in Southern Sudan were implemented by the Ministry of Health, Government of Southern Sudan (MoH-GoSS), with technical support from Malaria Consortium. In Kenya, surveys were implemented with the help of the Kenya Medical Research Institute - Wellcome Trust Research Programme (KEMRI-WTRP). I was involved in the planning, implementation and analyses stages in both Southern Sudan and Kenya. During both surveys I co-led a team of health workers in specific areas, ensuring the smooth running of the survey. Additionally, in Southern Sudan I was in charge of selecting the survey sites, designing data entry sheets and Personal Digital Assistant (PDA)

templates, pre survey staff data entry training, managing datasets in the field, double entry of data and data cleaning.

2.2.2 Survey context

Southern Sudan

Southern Sudan is now an independent country, having split from the north (the Republic of Sudan) in July 2011. The country has experienced extended periods of conflict for the last 50 years. This volatile setting has allowed very little disease control and, as such, the country is thought to be among those with the highest per-capita burden of NTDs in the world (Rumunu et al., 2009). The relative stability brought about by the signing of a peace agreement in 2005, has enabled the MoH-GoSS to be established and develop plans for the rebuilding of the health sector. Among the priorities is the control of NTDs and the government has now committed itself to integrated control using PCT, as recommended by the WHO. To target this intervention, however, it was necessary to identify which geographical areas require intervention. Technical and logistical support for the implementation of the integrated NTD control programme is being provided by Malaria Consortium.

The distribution of onchocerciasis in nearly all endemic countries in Africa has been comprehensively mapped with assistance from the APOC (Noma et al., 2002; Etya'ale, 2008). In Southern Sudan, trachoma surveys have also been undertaken in a number of areas (Ngondi et al., 2006; King et al., 2008; Kur et al., 2009). For LF, schistosomiasis and STH, by contrast, systematic prevalence data are not available, with previous studies in Southern Sudan having been few and

limited in scale (Homeida *et al.*, 1994; Magambo *et al.*, 1998; Deganello *et al.*, 2007). For these reasons, it was deemed necessary to conduct comprehensive surveys for schistosomiasis, STH infection, LF, and loiasis (due to the worm *Loa loa*) across the country to guide the design and integrated delivery of PCT packages to endemic areas.

Kenya

It is estimated that more than five million school-aged children are at risk of infection with one of the three major STH species. In response, the Government of Kenya has implemented a national School Health and Nutrition Programme (SHNP) which, with financial support from the World Bank and technical support from KEMRI-WTRP, aims to deworm at risk school children to reduce prevalence of heavy intensity infections of STH and schistosomiasis to below 1%, the threshold above which disease is considered to be a public health problem. Additionally, the government recently launched its National Malaria Strategy, which aims to target malaria interventions specific to local transmission dynamics. As part of these new initiatives, survey data were required to map and describe the distribution of helminth infection and malaria and to provide baseline data with which to measure the impact of subsequent interventions. The initial focus of these surveys was Coast province, which has historically been one of the most seriously affected areas in the country. Schools provide both an operationally and epidemiologically sensible unit through which to carry out surveillance of helminth infections in Kenya, due to high disease prevalence school-aged children and high school enrolment (Brooker *et al.*, 2009c).

2.2.3 Sample population and selection

Southern Sudan

The survey was based on an integrated NTD survey protocol developed by the MoH-GoSS and Malaria Consortium (MoH-GOSS, 2009), with support from the US Agency for International Development and technical input from the Centers for Disease Control and Prevention and Research Triangle Institute International. The protocol followed WHO recommendations for each of the NTDs, with slight modifications to improve feasibility in the challenging context of Southern Sudan.

The survey was conducted in 86 villages in Northern Bahr-el-Ghazal State, north-western Southern Sudan, from February to May 2009. In Southern Sudan, the first administrative unit is the state, followed by county (2nd) and payam (3rd). Northern Bahr-el-Ghazal State is divided into five counties and 18 payams, and has a population of approximately 1,580,695, which amounts to about 12% of the total population of Southern Sudan. The State experiences a single rainy season, typically between June and September. The population mainly consists of the Dinka ethnic group, who engage in nomadic cattle herding at riverside camps during the dry season and growing of millet and other varieties of grain in fixed settlements during the rainy season. Like most of Southern Sudan, this State is characterized by a lack of physical infrastructure and occasional insecurity, making the conduct of surveys particularly challenging.

In each payam, a two-stage, quasi-random sampling approach was employed. Initially, in order to maximise identification of high LF prevalence areas, selection of villages was based on anecdotal reports on the presence of lymphoedema and hydrocele, collected through interviews with payam administrative and medical staff. In accordance with WHO recommendations, sampling for LF in each payam was conducted until a maximum of 250 individuals had tested negative, which required visits to up to three villages to reach the required sample size. In addition, in those villages selected for LF, 32 village chiefs were interviewed regarding the presence of clinical manifestations of LF in their village, and a rapid assessment for *Loa loa* was conducted. The majority of villages surveyed for LF were also surveyed for schistosomiasis and STH. Due to problems with the supply of ICT kits, 13 villages that should have been surveyed for both LF and schistosomiasis/STH were instead only selected based on proximity to water and only surveyed for schistosomiasis/STH. When ICT kits were available, 13 different villages were selected for LF surveys, based on anecdotal reports of clinical manifestations. Individuals were excluded from the study if they had not lived in the payam for at least six months or did not provide informed consent.

For schistosomiasis and STH, the number of villages to be sampled in each payam was calculated according to the population size of each payam, whilst ensuring a minimum of two geographically well-separated sites were selected per payam. To guide selection of villages in addition to those already selected through our LF sampling strategy, a list of villages within areas of expected schistosomiasis risk was generated. This list was derived from an initial map of expected risk based on climatic and ecological information. However, due to the lack of a georeferenced village

database, a list of villages close to water bodies was compiled during interviews with payam administrative staff and a random selection was taken. If this selection procedure did not generate a sufficient number of villages, additional ones were chosen by using a randomly generated list.

STH infection was assumed to be geographically more homogeneously distributed than schistosomiasis and LF (Srividya *et al.*, 2002; Brooker *et al.*, 2004b; Brooker, 2007), and therefore selection of sites on the basis of LF and schistosomiasis ecology was considered sufficient to capture the inherent spatial heterogeneity of STH infection. Such a sampling approach was based on the operational requirement to identify whether a particular payam required MDA of PCT for a particular NTD, rather than to formally estimate the prevalence of species infection.

Kenya

The study took place in rural primary schools in the former districts (as of 1999) of Kilifi, Kwale and Malindi in Coast Province. These areas were the initial focus of the Government of Kenya's SHNP implemented in 2009. A two stage sampling design was employed: (i) schools (primary sampling units) were selected using probability proportional to size; (ii) within schools, a fixed number (n=100) of children (secondary sampling units) were randomly selected from each of these schools. Sample size calculations were based on the need to a) estimate the prevalence of infection in each district with adequate precision; and b) to have sufficient power to detect changes in infection prevalence due to deworming. The sampling frame was all public primary schools within a district. The number of schools required within the sample depends upon the

likely design effect (the change in sample variance using cluster sampling over simple random sampling) (Williams *et al.*, 2008) for each of the three parasitic infections of interest. Here it was assumed, in the absence of empirical data, the design effect for hookworm to be 1.5, *P. falciparum* 2 and *S. haematobium* 4. The final sample size was calculated to give adequate precision for estimating the most clustered infection, *S. haematobium*. Assuming that prevalence of *S. haematobium* is at least 10% it was estimated that in each of the three districts at least 2,014 children in 21 schools would be required. To account for potential school closures and non-compliance, 25 schools per district were selected, yielding a total of 7,500 children in 75 schools. This sample size was also sufficient to detect a 50% reduction in the prevalence of heavy *S. haematobium* infection in the post-intervention evaluation compared to baseline levels at 5% level of significance and 80% power to detect a 33% reduction in the prevalence of anaemia in the post-intervention evaluation compared to baseline levels at 5% level of significance.

Coordinates for each school/community were collected using handheld Global Positioning System devices (eTrex, Garmin International Inc., Kansas, U.S.A.) in both studies.

2.2.4 Survey methods

Southern Sudan

In each selected village, meetings were held with village elders to explain the nature and purpose of the survey. In those villages where LF surveys were conducted, the village chief was also interviewed regarding numbers of residents with clinical manifestations of LF, using the standard WHO questionnaire (WHO, 2000b). Initially, households were selected using the random walk

approach (Bennett et al., 1994); however, in some villages households were too dispersed for this method to be feasible, and registration of individuals subsequently occurred at a central location. For schistosomiasis and STH, children aged 5 to 15 years were invited to participate, and households were selected until a sample of 70 children was registered in each village. Selected children were given containers for stool and urine samples, and asked to drop the sample off at a central point, where the field laboratory had been established.

For LF, individuals aged 16 years and above were invited to participate, and households were selected until a total of 110 adults had been registered in each village. Those registered were requested to provide a finger-prick blood sample to be tested for circulating W. bancrofti antigen using an ICT (BinaxNOW® Filariasis, Inverness Medical). ICT kits were refrigerated whilst in storage in Nairobi, Juba and Aweil, according to guidelines, and were kept in cool boxes throughout fieldwork. In accordance with WHO guidelines (WHO, 2000b), if one or more tests had been positive in a sample of 100 individuals, then no further testing would have been undertaken. In practice, however, this scenario did not occur in any payam, and therefore a second site in the payam had to be selected, with a further 110 adults registered and requested to provide a finger-prick blood sample. In addition, those children registered for the schistosomiasis and STH survey in the selected village were also requested to provide a finger-prick blood sample so that a total of 180 individuals were registered for LF at the second site. If there were insufficient individuals in the second village to reach a total sample size of 250, then sampling continued in a neighbouring, third, village. In villages surveyed for LF, data on the presence of L. loa were collected from each adult registered for ICT testing using the WHO recommended Rapid Assessment Procedure for

Loiasis (RAPLOA) (TDR, 2002). This procedure consists of asking a sample of 80 adults per community three questions regarding the presence of worms in their eyes. Consistent with RAPLOA guidelines, children tested for LF were not interviewed for *L. loa*.

Parasitological examination of stool and urine samples was conducted in the field by a team of trained laboratory technicians. Faecal samples were examined in duplicate for *S. mansoni* and STH ova using the Kato-Katz method within an hour of preparation to avoid the clearing of hookworm eggs. Urine samples were tested for haematuria using Hemastix® reagent strips (Bayer Corporation), with test positive urine samples subsequently examined using urine filtration (Bergquist *et al.*, 2009). Most technicians were skilled in conducting all of the survey tasks and were regularly rotated between: registration of individuals for the survey, preparation of stool and urine slides for microscopic examination, collecting finger prick blood samples for ICT kits and carrying out microscopic examination of slides.

Kenya

Selected schools were visited one week prior to the survey date to have the purpose of the survey explained to the head teacher and school committee, and permission was sought at the school-level. Six staff travelling in a single vehicle visited each school: one supervisor; three laboratory technicians; one laboratory assistant; plus one driver. An initial meeting was held with the head teacher and then pupils in classes 2-6, who were willing to participate in the study and were between the age of 5-16, were registered to take part in the survey. Pupils unwilling to participate, or who were outside this age range, were excluded from the study.

A series of questions were asked of each randomly selected participating child including: age, fever on the day of the survey, use of mosquito nets treated with insecticide, source of potable water in the homestead, other basic household assets indicators, distance of house from school and whether other siblings attend the same school. A school questionnaire was also administered to the head teacher to collect information on enrollment, water and sanitation facilities, and previous involvement in deworming and school health activities.

On the day of enrollment to the study, pots for stool and urine samples were distributed to all selected children who were asked for a sample. Each child was also asked to provide a finger-prick blood sample to conduct an Optimal* rapid diagnostic test for malaria parasites. Additionally, blood was used to prepare a thick and thin blood smear for slide examination of malaria parasites. A child was classified as malaria positive if either the RDT or slide suggested a positive result.

2.2.5 Ethical considerations

Southern Sudan

The study protocol received ethical approval from the Directorate of Research, Planning and Health System Development, MoH-GoSS, and from the Ethics Committee of the London School of Hygiene and Tropical Medicine, UK (reference # 5500). Clearance to conduct the surveys was obtained from the State MoH, followed by County Health Departments. The study was explained

to each member of the selected households. The household heads were asked to provide written consent for the entire household to participate in the study. The study was then explained to each household inhabitant who met the inclusion criteria, and s/he was asked to consent verbally to participating in the study; only those who did were registered and requested to provide samples. Due to the large number of individuals that were surveyed, it was considered impractical to obtain written consent from each study participant, and the institutional review boards approved this procedure. To document verbal consent, the name of each individual who provided verbal consent was recorded, along with the test results for their samples. Those individuals who did not consent were not registered nor examined. Individuals who tested positive for schistosomiasis or STH infection were treated with praziquantel or albendazole respectively, according to WHO guidelines. The four individuals positive for *W. bancrofti* antigen were not treated on site, but informed of their infection and its possible implications, after which they were referred for treatment to the nearest health facility. On the basis of advice from the MoH-GoSS, this approach was considered more appropriate, as repeated treatment will be required to kill microfilaria and thus prevent transmission.

Kenya

The study protocol received ethical approval from the Kenya Medical Research Institute and National Ethics Review Committee (reference # 1407 and 1596). Head teachers were briefed about the survey and were provided with an information sheet detailing the survey procedures and asking for their permission to have their school involved in the survey. The head teachers were also asked to inform the students, parents and the school committee members about the survey

and obtain their approval for the study. Parents/guardians were free to refuse participation of their child should they wish. On the survey day, all children in the school were informed about the details of the survey and were told that participation was voluntary. Individual written parental consent was not sought since the survey was conducted under the auspices of the Division of Malaria Control, Ministry of Public Health and Sanitation, which has the legal mandate to conduct routine malaria surveillance, and because only routine diagnostic procedures were undertaken. As examination of stool and urine slides did not take place in the field, those children found positive for either schistosomiasis or STH did not receive treatment but all schools received mass treatment as part of the 2009 national school deworming programme. Any children with a positive RDT result had their temperature taken, and, if this was >37.5°C, arrangements were made for immediate referral to the local health facility.

2.2.5 Data analysis

For both surveys, data were double-entered into Microsoft® Excel®. Data checks and first entry were conducted at the end of each survey day. Second entry was mostly conducted during the period of field work and completed afterwards. Range and consistency checks were conducted for all variables. Maps of infection prevalence were developed using ArcGIS 9.2 (ESRI, California, U.S.A.).

2.3 RESULTS

Southern Sudan

In total, 4,904 children and 4,834 adults from 86 villages across Northern Bahr-el-Ghazal State were registered to take part in the survey. Only Agoga payam, in the north of the state, was inaccessible by vehicle during the study period. For schistosomiasis and STH infection, 73 villages were surveyed and a total of 4,450 stool samples and 4,597 urine samples were examined. For LF, 5,254 blood samples from 43 villages were tested. Two of the sites surveyed for LF had insufficient inhabitants to make up the required sample of 110 per village, thus requiring inclusion of individuals from the neighbouring village. These data were merged before analysis due to the close proximity of the sites. Of the children that provided either a stool or urine sample, the mean age was 8.9 years (inter-quartile range (IQR): 7 - 11 years) and 50% were male. Of the adults that provided a blood sample, the mean age was 36.7 years (IQR: 25 - 45 years) and 36.6% were male. The bias towards females was mainly due to the fact that men spent the day fishing outside the village.

The overall prevalence of *S. haematobium* was 3.0% (0 – 65.6% by village) and *S. mansoni* was 0.2% (0 – 4.2% by village) (Figure 2.1). Although state-wide levels of infection were low, there was marked geographical variation, with prevalence of *S. haematobium* >20% in some villages along the Loll river. The most common STH infection found in the state was hookworm: the overall prevalence was 4.9% (0 – 70% by village) (Figure 2.1). Hookworm prevalence showed a strong geographical pattern, exceeding 20% in the south of the state. *A. lumbricoides* was only detected in one individual and no individuals were found to be positive for *T. trichiura*.

S. haematobium prevalence in Northern Bahr-El-Ghazal
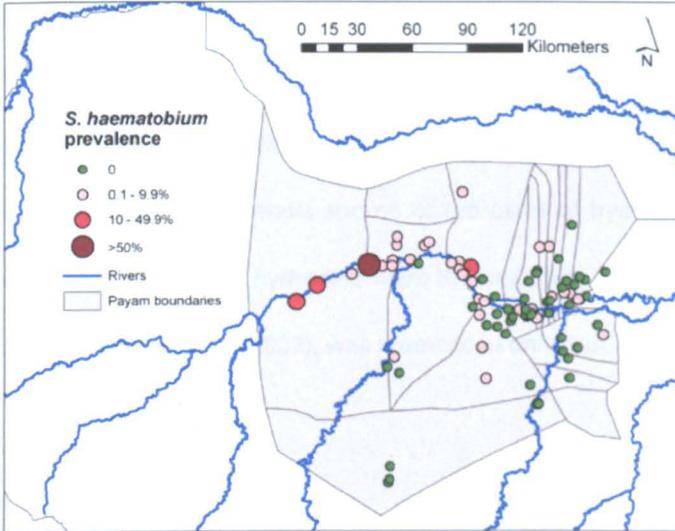
Map of Africa showing
the location of Sudan

Hookworm prevalence in Northern Bahr-El-Ghazal
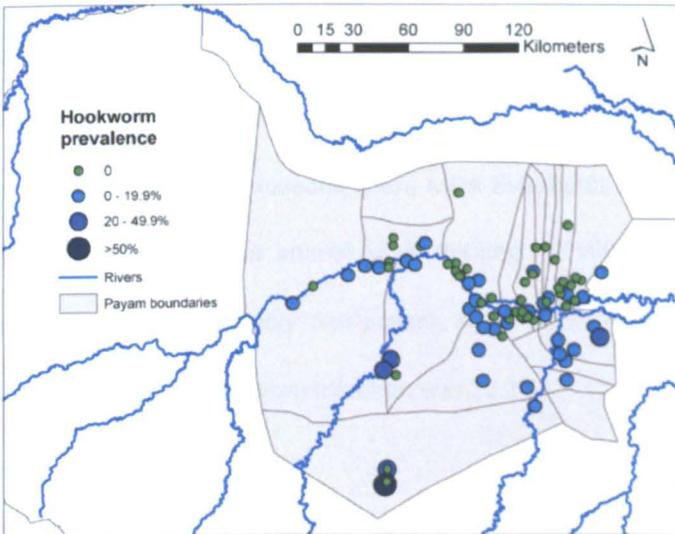
Map of Sudan showing
the location of Northern
Bahr-el-Ghazal

Figure 2.1 The distribution of *S. haematobium* and hookworm in Northern Bahr-el-Ghazal State

For LF, no payams had an antigenaemia prevalence of above 1%; overall only four ICT positives were detected, two each in Aweil Centre and Aweil East payam. Questionnaires on the presence of

clinical manifestations of LF in the community were administered to village chiefs in 31 of the 43 villages surveyed with ICTs. In 84% (26) and 71% (22) of villages, respondents reported having seen villagers with elephantiasis or hydrocele, respectively. Questionnaire data were only available for two of the four villages with positive ICT results, in both of which respondents reported having seen one case of elephantiasis and no or ten cases of hydrocele. All other villages with reported cases of elephantiasis or hydrocele were ICT negative. History of eye worm, calculated according to WHO guidelines (TDR, 2002), was reported in only four individuals from two villages.

In terms of treatment, WHO guidelines recommend that in communities with schistosomiasis prevalence of ≥10% and <50%, school-aged children and high risk groups of adults should be treated with praziquantel once every two years. In communities where prevalence is ≥50%, the same groups should be treated once a year (WHO, 2006c). In Northern Bahr-el-Ghazal State, *S. haematobium* infection exceeded these MDA thresholds in only four of the survey communities, one of which qualifies for annual MDA. Pooling the village prevalence data at the payam level would have meant that only one payam, Ayat, would have qualified for MDA, because overall prevalence of *S. haematobium* infection was 20.2%.

For STH, WHO recommends delivery of MDA with either albendazole or mebendazole once a year to pre-school and school-aged children, as well as to pregnant women and high risk groups of adults, where cumulative prevalence of STH is ≥20% and <50%, and twice a year where prevalence is ≥50% (WHO, 2006c). In Northern Bahr-el-Ghazal State only five of the survey communities

exceeded the MDA threshold for STH, due to the presence of hookworm, with only one community falling into the biannual MDA category. Pooling of survey data at the payam level would result in two payams, Aroyo and Awoda, exceeding the MDA intervention threshold, with a mean hookworm prevalence of 26.9% and 56.5%, respectively. Payams that did not qualify for MDA for either *S. mansoni* or STH contained one village with a prevalence of *S. haematobium* of ≥10%.

Kenya

In total, 6,667 children from 65 schools were included in the survey, with a mean of 102.6 children surveyed per school (range: 82 – 111). The mean age was 12.18 years (IQR: 10 - 14 years) and 50.7% were male. Overall prevalence of *S. haematobium* was 13% (range: 0 – 77%), cumulative STH 29.8% (range: 0.9 – 75%) and *P. falciparum* 4% (range: 0 - 29%). Of the STH, hookworm was the most prevalent (22.4%) followed by *T. trichiura* (11.1%) and *A. lumbricoides* (2.6%). Infection with an STH species appeared to be more prevalent towards coastal areas (Figure 2.2). Infection with *S. haematobium* and *P. falciparum* both appeared to be more heterogeneous, with *P. falciparum* more prevalent in the southern half of the province.

In terms of treatment recommendations, 20 schools had a prevalence of *S. haematobium* ≥10% and therefore qualified for biennial mass treatment with praziquantel. Of those, six schools had a prevalence ≥50% and therefore qualified for mass treatment with praziquantel once a year. For STH, 35 schools had a cumulative prevalence ≥20% and therefore qualify for mass treatment with

albendazole once a year. Of those, 17 had a prevalence of ≥50% and therefore qualify for mass treatment twice a year. For malaria, there are as yet no clear guidelines for thresholds at which different interventions should be introduced. Hay *et al.* (2008), suggest that in areas where prevalence of *P. falciparum* infection is <5%, targeted use of ACT, ITN and insecticides should be used. Where prevalence is ≥5%, there should be universal coverage of ITNs, IRS and ACT. According to these recommended thresholds, 17 schools had a prevalence of ≥5%.

If data are pooled at the district level, only Kwale district had a prevalence of *S. haematobium* of ≥10%. Kilifi and Kwale districts had a cumulative STH prevalence of ≥20% and therefore qualify for MDA of albendazole. For malaria, only Kwale had a prevalence of ≥5%. Districts that did not exceed the intervention thresholds for these parasites contained six schools with a *S. haematobium* prevalence of ≥10% and one school with a prevalence of ≥50%, two schools with a cumulative STH prevalence of ≥20% and five schools with a *P. falciparum* prevalence of ≥5%.

Map of Africa showing the location of Kenya

Map of Kenya, showing the location of the survey districts

S. haematobium prevalence
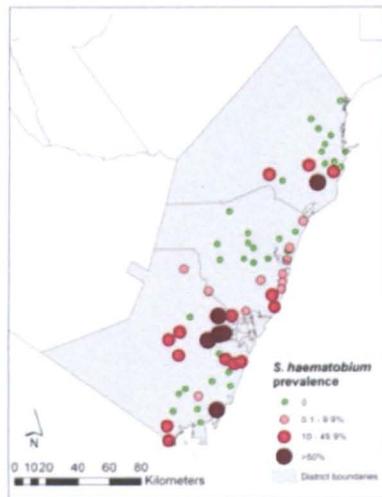
Cumulative STH prevalence
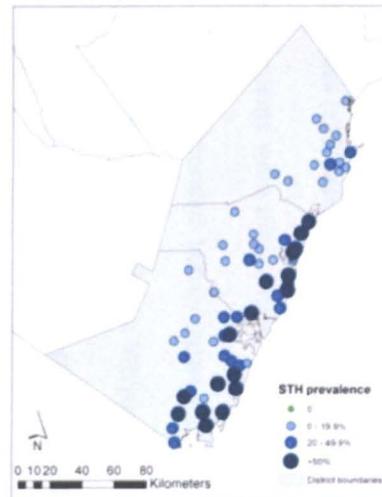
P. falciparum prevalence
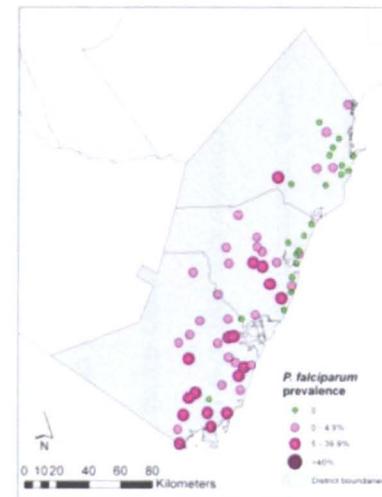
Figure 2.2 The distribution of *S. haematobium*, STH and *P. falciparum* in Coast Province, Kenya.

2.3.1 Practical lessons learnt

Designing and implementing an integrated survey added a complex, although not insurmountable, operational layer. Integrated surveys in Kenya were easier to implement due to the fact that the same schools, and individuals within those schools, were included in surveys for all parasites. The large number of survey schools did, however, incur considerable cost and required separate surveys teams for the different districts in order to complete the surveys during the school term time. An investigation into the balance between cost and performance of different sample sizes would be beneficial. In Southern Sudan, it was not possible to screen the same study participants for all the parasites surveyed for as, according to current recommendations, individuals of differing ages are required for the different diseases (>16 years for LF and 5-15 years for STH and schistosomiasis). Additionally, sample size requirements were different for the different diseases. Furthermore, in order to stay as close to the recommended survey approaches as possible, different combinations of diseases were surveyed for in different villages. Whilst this did not lead to any errors being made, it added further operational complexity. The challenging context of Southern Sudan also made it difficult to maintain the cold chain required for the ICT kits, making it difficult to be absolutely certain that every kit was in optimal condition.

2.4 DISCUSSION

A sound understanding of NTD distribution and prevalence is an essential prerequisite for cost-effective control, with each national programme needing to be tailored to its specific context. MDA should be targeted to those areas and populations in greatest need and programme managers hence require information on populations-at-risk and numbers to be treated to estimate the funding needed to deliver the intervention. Here results are presented from two recent attempts to carry out integrated NTD surveys in Southern Sudan and Kenya.

The survey results illustrate some important features of these diseases that are worth highlighting. In Southern Sudan and Kenya, schistosomiasis and malaria appeared to be more focal in their distribution than STH. As discussed in the introduction, this is most likely due to the fact that *S. mansoni* and *P. falciparum* rely on an intermediate snail host and mosquito vector respectively, which acts to constrain transmission to areas conducive to the overlap of parasite and definitive and intermediate hosts (Shope, 1999; Brooker, 2007; Sturrock *et al.*, 2010). These differences in the scales over which diseases cluster should influence the scales over which decisions on intervention are made. This is illustrated in the results of the surveys reported here. In both Southern Sudan and Kenya, it is evident that there are a number of communities that qualify for MDA with praziquantel based on schistosomiasis prevalence that would miss treatment if results are aggregated at the payam/district level. Likewise, in Kenya, there are likely to be a number of schools with a prevalence of *P. falciparum* of ≥5% that would miss any intervention if results were pooled at the district level. This illustrates the fact that foci of transmission of these diseases can occur within low prevalence districts. In contrast, if STH data were aggregated at payam/district level, all the communities surveyed that qualified for STH treatment in Southern Sudan would

have been treated and in Kenya only two schools would miss treatment. This suggests that prevalence of STH is more homogeneous within payam/district than *S. mansoni* or *P. falciparum*, and that using the payam/district as an IU is epidemiologically justified. Despite the importance of spatial heterogeneity on the design of control programmes, and surveys, this area of research has largely been ignored and requires further attention.

The survey designs outlined here, represent a first step in combining current WHO guidelines for individual parasite surveys into a single integrated design in two different settings. In the process, a number of limitations were identified. In Southern Sudan, implementation was largely affected by the challenges brought about by operating in a post-conflict setting. First, the lack of up-to-date census data and a georeferenced village database - due to longstanding civil war - meant that villages could not be selected entirely at random from within areas identified to be at risk of schistosomiasis. Instead, local knowledge had to be used to identify sites where schistosomiasis (as well as clinical manifestations of LF) had been reported from and that were accessible. Second, the use of purposive sampling may have resulted in slightly higher prevalence estimates when data were pooled at the payam level. Third, in the more dispersed villages it was not feasible to implement a random walk selection procedure and a convenience sample was selected at a central point. This approach may have introduced sampling bias through (i) individuals with potential clinical signs of disease being more likely to attend because of the offer of diagnosis and treatment and (ii) ill individuals unable to attend (Levy and Lemeshow, 1999). However, village leaders were used to mobilise individuals, and although bias may have been introduced, it is unlikely to have altered the overall treatment classifications. Recent experience of a filariasis treatment coverage survey in Haiti found little difference between coverage estimates obtained

through a convenience sample of houses near distribution points and a cluster survey (Mathieu *et al.*, 2003). Fourth, the age group surveyed for STH and schistosomiasis included children between the ages of 5-15 years, whereas children between the ages of 10-15 years are normally targeted because prevalence typically peaks in this age group. The decision to use a wider age band was made to maximise the number of children surveyed in a village so as to avoid having to visit a second village. During preliminary visits it was found that villages are often small and would not contain sufficient numbers of 10-15 year olds to complete sampling. Whilst these limitations may have led to some areas being falsely classified as requiring treatment, it should be noted that no survey methods provide perfect results and the approach discussed here provides a balance between minimising survey effort and maximising the probability correctly treating infected populations.

In Kenya, the surveys were designed to provide good prevalence estimates with specified confidence. In terms of providing information on whether to intervene, however, such a survey design has a number of disadvantages. Firstly, as decisions on control of STH are regularly made at the district level, it is likely that 20 schools per district is an excessively large sample size with which to classify a district according to treatment prevalence thresholds. A reduction in the number of schools is likely to result in only a small loss of performance but would allow considerable cost savings. Further work is required to explore the balance between sampling effort and the ability to classify districts according to treatment thresholds. Secondly, in the case of *S. haematobium,* decisions on control are often taken at the school level due to the focal nature of the disease. Mapping prevalence at the district level, however, does not allow this level of resolution and survey methods that aim to provide a rapid and cost-effective means to identify

schools that qualify for treatment are required. Questionnaires administered to pupils which enquire about symptoms of disease have been used for this purpose in a number of countries (Lengeler *et al.*, 2002). Alternatively, LQAS could be combined with the use of rapid diagnostic tests such as haematuria reagent strips that provide a cheaper alternative to parasitological examination of eggs in urine. Thirdly, it is possible that the use of schools may have introduced sampling bias as children suffering from the effects of STH, schistosomiasis or malaria on the day of the survey may have been unable to attend school and would therefore have been missed. Fourthly, the usefulness of schools as indicators of local infection prevalence is somewhat dependent on the size of the catchment area. Some children reported walking several kilometres to attend school.

Whilst discussing the limitations of these surveys is valuable, it is also important to highlight the benefits of an integrated approach to NTD mapping. In both settings, it is unlikely that funding would have been available to carry out separate surveys for all the diseases covered by the integrated approach. Furthermore, an integrated approach allows information on all diseases to be collected at the same time, which is essential for planning integrated control packages.

The two studies reported here identify a number of issues that require further investigation in order to assess the performance and feasibility of integrated NTD surveys. First, it is apparent that a better understanding of the spatial heterogeneity of different NTDs is required to help inform the optimal number of schools/communities that need to be surveyed to reach a decision about MDA. Second, more detailed information is required on the cost implications of different survey approaches particularly when taking account of the cost of drugs (and their delivery). The next

chapter begins this process by investigating the spatial heterogeneity of surveys for STH and the cost-effectiveness of alternative survey designs. Subsequent chapters will extend this work to schistosomiasis and malaria.

CHAPTER 3 | OPTIMAL SURVEY DESIGNS FOR TARGETING CHEMOTHERAPY AGAINST SOIL-TRANSMITTED HELMINTHS: EFFECT OF SPATIAL HETEROGENEITY AND COST-EFFECTIVENESS OF SAMPLING

3.1 INTRODUCTION

Building on results from chapter 2, this chapter investigates optimal survey strategies for STH. For the mapping of STH, WHO recommends that school-aged children aged 7-14 years are sampled, with 50 children selected per school or community (Montresor *et al.*, 1998). In each ecological zone, it is recommended that five schools are randomly selected. However, the empirical basis for these recommendations is unclear and there is a need to investigate optimal survey designs to target MDA for STH.

To date, surprisingly, no study has investigated the utility of rapid survey procedures to support the control of STH. This may be partially explained by two factors. First, benzimidazoles used for treatment cost approximately US$0.02 per person treated and are therefore considered cheap enough to distribute uniformly throughout countries with no evidence-based targeting. In reality, however, many national governments still do not have sufficient resources to support the large scale delivery of drugs required for comprehensive treatment strategies. Second, STHs have been

assumed to be geographically homogeneous, with similar infection levels occurring over large distances (Brooker *et al.*, 2004b). Spatial heterogeneity has important consequences for surveys and control as it determines the resolution at which surveys and interventions should be carried out. Diseases that are very widespread and have similar levels of infection over large areas require fewer survey points than more focal diseases that require higher resolution data to avoid missing foci of infection. In practice, however, there are few studies that have quantitatively explored this issue for STH at a scale of operational relevance.

Defining an optimal sampling scheme for targeting STH control requires an understanding of the following issues: (i) the degree of spatial heterogeneity of STH infection; (ii) the financial and human cost of conducting epidemiological surveys for STH; (iii) the geographical framework within which public health decision-making is organized through community, sub-district and district levels; and (iv) the financial and public health consequences of inappropriate control decisions on the need for mass treatment. The aim of this study is to quantify the spatial heterogeneity of STH infection in a range of transmission settings in eastern Africa and use this information to investigate the accuracy and cost implications of alternative sampling strategies in order to classify intervention units according to treatment strategy.

3.2 METHODS

3.2.1 Overview

The spatial heterogeneity of STH species was characterized using geostatistical analysis of data on the prevalence of infection in school children from four countries in eastern Africa, allowing comparison of results over a range of transmission and ecological settings. Using the example of Kenya, these spatial characteristics were used to parameterize simulation analyses that explored the implications of survey designs for enumerating district-level (2nd administrative level) infection status and informing treatment strategies. Alternative sampling schemes were evaluated in terms of both their reliability in classifying districts according to appropriate treatment strategy and their cost implications when considering the combined cost of survey and treatment. Kenya was selected as an example due to the availability of (i) a national, georeferenced school database and (ii) detailed, standardised survey data from Kenya and from bordering areas supporting empirical estimates of spatial heterogeneity.

3.2.2 Empirical data sources

The data used to quantify the spatial heterogeneity of STH species were estimates of infection prevalence from single national or sub-national surveys that applied standardized methodologies (i.e. examination of school aged children using Kato-Katz examination of stool samples) (WHO, 1991) for coastal Kenya, Uganda, northwest Tanzania, and Zambia (Table 3.1 and Figure 3.1). These surveys represent some of the most geographically comprehensive survey data for STH in sub-Saharan Africa.

Table 3.1 Summary of data on STH infection used in the geostatistical analysis of spatial heterogeneity.

Country	Number of schools	Mean number examined per school	Mean age examined (years)	Median prevalence (%) (range by school)			Reference
				Hookworm	*A. lumbricoides*	*T. trichiura*	
Kenya	64	104.3	12.2	16.7 (0-69.2)	0 (0-23.4)	4.7 (0-54.4)	(Brooker unpublished data)
Tanzania	197	75.9	14.03	43.3 (1.6-93.3)	0 (0-40)	0 (0-43.3)	(Clements *et al.*, 2006a)
Uganda	197	68.9	10.21	52.4 (0-90)	0 (0-65.8)	0 (0-40)	(Kabatereine *et al.*, 2005)
Zambia	79	59.8	13.06	11.1 (0-86.7)	0 (0-33.3)	0 (0-21.3)	(Mwanza *et al.*, unpublished data)
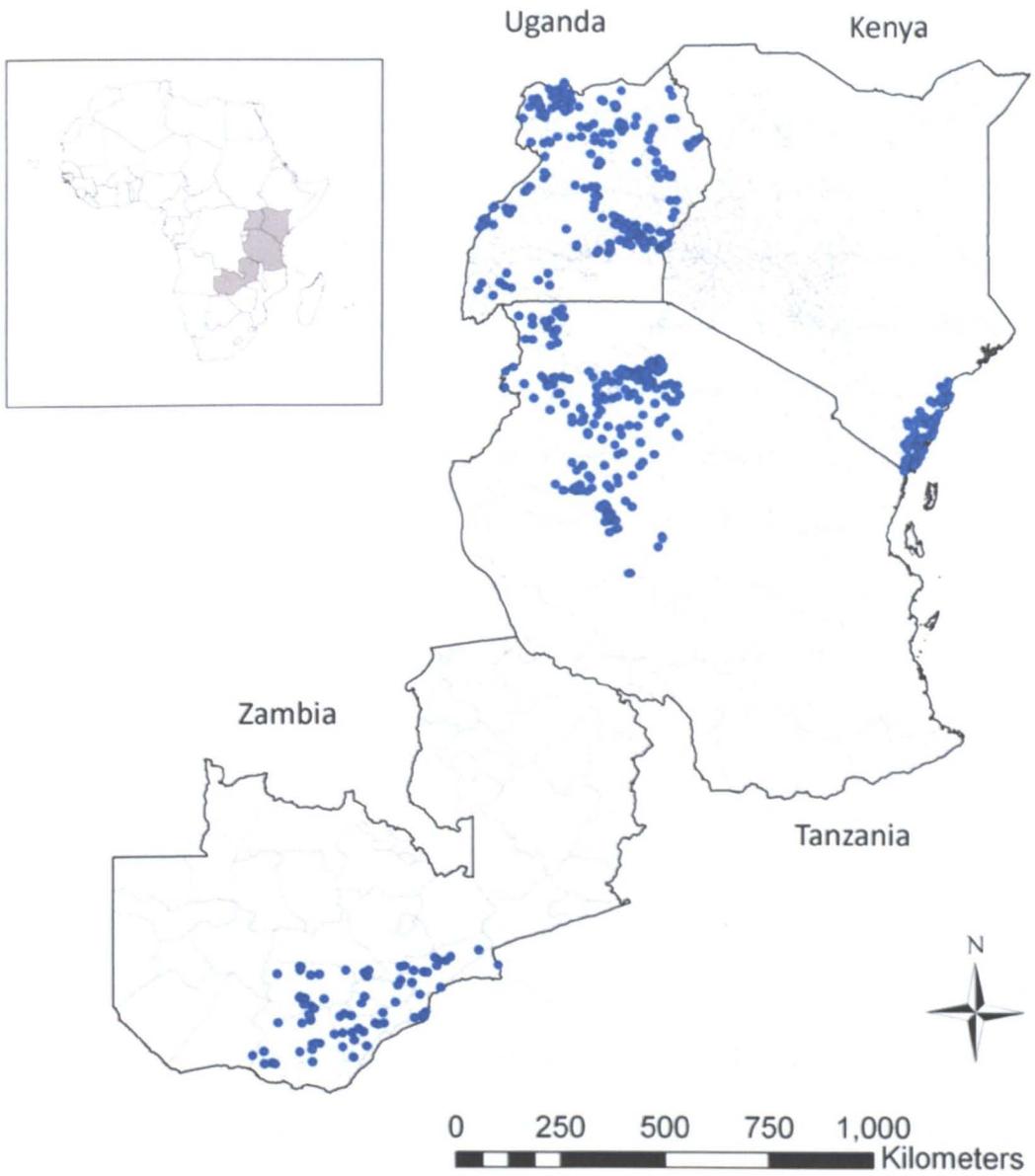
Figure 3.1 Locations of the countries included in the study (inset map) and locations of schools from the datasets included in the study (main map).

3.2.3 Geostatistical analyses

Spatial heterogeneity of infection prevalence was investigated using semi-variogram analysis. A semi-variogram characterises the spatial autocorrelation structure of a variable by defining semi-variance (a measure of expected dissimilarity between a given pair of observations made at different locations in space) as a function of lag (the distance separating the observation locations). A semi-variogram can be estimated from survey data by measuring the mean squared difference of pairs of observations that are separated by the same distance (termed lag) (Oliver *et al.*, 1992; Diggle and Ribeiro Jnr, 2007). Information about the spatial autocorrelation structure and the distance over which this occurs can be inferred from the shape of the semi-variogram. If spatial autocorrelation is evident, semi-variance typically rises with distance, eventually plateauing to a maximum value termed the sill. The separation distance at which the sill is reached is termed the range and represents the maximum distance over which values are autocorrelated, with larger separation distances implying spatial independence. The value where the semi-variogram intercepts the *y*-axis is called the nugget variance, and represents measurement error or spatial autocorrelation occurring over distances smaller than those represented in the data (Cressie, 2000). A uniformly flat semi-variogram is indicative of an absence of spatial autocorrelation, with even closely located points varying independently. An 'unbounded' semi-variogram that rises continually without reaching a plateau is indicative of an underlying trend: spatial autocorrelation operating over lags substantially larger than the study region.

The presence of a large-scale spatial trend hampers variogram analysis by obscuring the influence of smaller-scale heterogeneity. Where large scale trends were detected via inspection of raw prevalence variograms, data were de-trended using logistic regression models that predicted

prevalence as a function of survey location and land surface temperature, an established determinant of large-scale distribution of STH infection (Brooker *et al.*, 2006b). The resultant normally distributed Pearson residuals were used to estimate the semi-variogram. In the remaining countries where no evidence of large-scale trends were detected, due to the skewed nature of the data, a logistic transformation was used before semi-variogram analysis, $y = (\log((d+0.01)/(1-(d+0.01))))$, where d is the raw prevalence data and y denotes the transformed variable that was approximately normally distributed. In estimating semi-variograms, the maximum lag distance was set to half the maximum inter-point distance and equally sized distance bands containing at least 30 pairs were used. Semi-variograms were fitted using weighted least squares fits of exponential, spherical and gaussian models and examined by visual inspection (Diggle and Ribeiro Jnr, 2007). Analyses were carried out using the GeoR package in R 2.7.1 (Ribeiro Jnr and Diggle, 2001). Semi-variograms were generated for each country separately, except for Tanzania and Uganda which represented contiguous areas and were therefore analysed initially as a single dataset. Due to large regional differences in prevalence of *A. lumbricoides* and *T. trichiura*, the contiguous Uganda and Tanzania dataset was split above and below 2° south before semi-variogram analysis of these parasites was carried out.

3.2.4 Conditional simulation and cost analysis: case study of Kenya

The exploration of spatial heterogeneity enabled the generation of a psuedo-dataset that had the same spatial and variance characteristics as those expected in the field. Simulating a completely enumerated dataset allowed different sampling strategies to be evaluated against a realistic 'gold standard'. To achieve such a gold standard, conditional simulation was used, which utilizes parameters arising from the semi-variogram analysis to generate a range of different scenarios (or

realisations) that reproduce the global characteristics of the source data in terms of the frequency distribution of input data values and the resultant semi-variograms. The process works as follows: a prediction location is randomly selected and a mean and variance are predicted using kriging (a spatial interpolation method that uses the semi-variogram to predict values based on data from known locations). Using the cumulative normal distribution with this predicted mean and variance, a random value is selected and this is the assigned prevalence value at that location. The procedure continues by selecting another prediction location and repeating the process until all locations have been visited. This then represents one realisation. As the final set of prevalence values in a realisation is dependent on the order of selection of prediction locations and the values assigned at each location, different realisations are unique in terms of locations of clusters and overall prevalence (Goovaerts, 1997). Data were simulated for all government mixed primary schools in Coast, Nyanza and Western provinces - the most populous provinces of Kenya and where STH are most prevalent (Brooker *et al.*, 2009b) - using the Kenya data and semi-variograms. Information on the schools and their location was obtained from the Ministry of Education school database (Figure 3.2). In total, data were simulated for 1,125 schools in seven districts in Coastal province, 2,046 schools in eight districts in Western province and 3,728 schools in twelve districts in Nyanza province. To allow sampling designs to be tested on contrasting scenarios, data were also conditionally simulated for schools in Western and Nyanza provinces using data and variograms from neighbouring Uganda.

Figure 3.2 Map of Kenya showing province boundaries (inset map) and locations of all 6,899 mixed public primary schools in Western, Nyanza and Coast provinces (main map). Note that the size and position of Western and Nyanza provinces are not in relation to Coast.

For each STH species and in each region, 1000 realisations were conditionally simulated. To generate estimates of cumulative STH prevalence (p, prevalence of any one of the three parasites) at each school, complete independence in the probability of co-infection was assumed using the following formula: $p = H + A + T - (HA) - (HT) - (AT) + (HAT)$ (Booth and Bundy, 1995), where H was the proportion infected with hookworm, A the proportion infected with A. lumbricoides and T the proportion infected with T. trichiura. A sensitivity analysis was undertaken to test this assumption of complete independence. For each of the 537 schools reported in Table 3.1, the expected prevalence of co-infection with different species combinations was calculated assuming complete independence and compared to the observed prevalences of co-infection. The observed

prevalence of co-infection for each species combination was plotted against the expected prevalence. A regression line was then fitted through this scatter plot so that observed co-infection could be estimated from expected. Next, observed probabilities of co-infection were used to estimate cumulative STH prevalence for each of the 6,899 simulated schools in Kenya and the implications for the performance of alternative sampling strategies explored.

Using these simulated data, and assuming a population of 500 at each school (a conservative estimate based on available data from Kenya Ministry of Education which suggest 420 children per primary school), the following sampling strategies were considered: a random selection of 1, 2, 3, 4, 5, 6 or 10 schools per district with a random selection of 10, 15, 20, 30, 40 or 60 children per school. For each sampling strategy, on each realisation, district prevalence was calculated by dividing the total number of positives per district by the number surveyed per district (N/n). The district was then classified according to WHO endemicity classes (<20% - low, ≥20% to <50% - medium, ≥50% - high) (WHO, 2006c). These classifications were compared to the 'true' endemicity class of the districts in each realization (i.e. the prevalence of infection in the district based on the fully enumerated simulated data), and the total proportion of districts correctly classified was calculated. Gross classification errors (i.e. high prevalence districts being classified as low prevalence districts and *vice-versa*) were also calculated. Districts ranged in size within each province: the median area of districts in Coast province was 7,861 kms^2 (range 232 - 38,701 kms^2), in Nyanza province 959 kms^2 (581 – 1,994 kms^2) and in Western province 937 kms^2 (556 – 2,058 kms^2). All the above simulations were carried out using bespoke scripts written in R 2.7.1 (R Development Core Team, 2008).

The cost of each sampling strategy was estimated using itemized costs collected for the Kenya survey. Following this approach, items were divided into staff, capital and consumables. Only the financial cost of the survey was estimated. Unit costs used in the costing are presented in Table 3.2 and were divided into fixed (irrespective of number of schools or children) and variable costs (which were dependent on the number of days and children). In terms of staff, it was assumed that one supervisor, one technician and one cleaner were required per day, irrespective of the number of children sampled (category 2). Where 31 – 59 children were sampled per school, an extra technician was included and if 60 or more children were sampled two further technicians were included (category 3). The remaining consumable costs were either dependent on the number of days (category 4) or the number of children (category 5). An average travel distance of 75 kms per day was assumed and a 10% contingency allowance was also included. Based on recent field experience in Kenya, it was also assumed that one school could be surveyed per day. Capital costs (category 1) were annuitized over the useful life of each item using a discount rate of 3%, consistent with the recommendations of the World Bank (World Bank, 1993). Such annuitization enables an equivalent annual cost to be estimated and reflects the value-in-use of capital items, rather than reflecting when the item was purchased (Walker and Kumaranayake, 2002). Vehicle running costs only included maintenance and insurance. Costs were estimated in local currency and their current values were converted into equivalent US$ using 1st September 2008 exchange rates of Kenyan Shillings 70.25 to US$1 and GBP 0.55 to US$1 (www.oanda.com/convert/classic).

In addition, the cost of treatment and delivery was calculated using two recent estimates of $0.15 and $0.39 per delivery round per child (Brooker *et al.*, 2008b; Hall *et al.*, 2009). Treatment was considered over one and five year periods. The total cost of each sampling strategy was therefore

estimated as the cost of the survey plus the cost of the MDA that would be carried out based on results of that survey. By including both survey and treatment costs and the proportion of districts correctly classified in the cost analysis, it is possible to include the cost of misclassification. To investigate cost-effectiveness of each sampling strategy, the cost per district correctly classified was calculated by dividing the average total cost of each sampling strategy across the 1000 realisations by the average number of districts correctly classified per realisation. For the purpose of this study cost-effectiveness of a sampling strategy is defined as the cost per district correctly classified.

Table 3.2 Itemized costs of conducting school surveys of STH infections in Kenya. The number of units required for 1 district using 4 schools are shown with ranges according to 10 – 60 children per school.

Cost type	Unit	Unit cost (US$)	Units required
Equipment	Microscopes	367.34	2
	Stool sieves	23.15	6
	Slide boxes	3.09	1
	Jerry cans	1.44	2
	Pencils	0.96	5
	Tally counters	8.62	4
	Buckets	0.96	2
	Wash basins	0.96	2
Salaries (Fixed)[1]	Supervisor	28.57	4
	Technician	14.29	4
	Cleaner	7.14	4
Salaries (Variable)[2]	Technician	14.29	0 - 8
Consumables (Fixed)[1]	Disposable gloves	5.00	0.8
	Bin bags	5.71	0.8
	Liquid soap	7.14	0.8
	Paper towels	2.14	4
Consumables (Variable)[2]	Kato-Katz kits	0.3	45 - 270
	Stool pots	0.05	45 - 270
	Wooden spatula	0.03	45 - 270
	Microscope slides	0.05	45 - 270
	Questionnaires	0.06	45 - 270
	Marker pens	4.00	0.8
	Biros	2.00	0.8
Transport[3]	Fuel and maintenance (per km)	1.43	300

[1] Fixed cost per day
[2] Variable cost, dependent on number of children
[3] Assumes an average distance of 75km per day

3.3 RESULTS

Prevalence data were available for 537 schools including some 39,924 children. The median prevalence of hookworm ranged from 11.1 to 52.4% between countries, while the median prevalence of *A. lumbricoides* across all countries was 0% and *T. trichiura* ranged from 0 to 4.7% (Table 3.1).

Figure 3.3 presents species-specific semi-variograms for each study region and shows distinct differences in the degree of spatial heterogeneity for hookworm compared to *A. lumbricoides* and *T. trichiura*. Specifically, the semi-variograms for hookworm indicate spatial autocorrelation with fitted range parameters between approximately 95 and 166 km. By contrast, the semi-variograms for *A. lumbricoides* revealed either no spatial autocorrelation or spatial autocorrelation with shorter fitted range parameters between 36 and 92 km. Similarly, the semi-variograms for *T. trichiura* only indicated spatial autocorrelation in three of the datasets, with ranges between 44 and 46 km, whilst other datasets showed little evidence of spatial autocorrelation. These results indicate that in eastern Africa (i) there is a consistency in the scale over which species-specific spatial autocorrelation occurs and (ii) that spatial autocorrelation in hookworm prevalence occurs over much larger distances than that for *A. lumbricoides* and *T. trichiura*.

In Kenya, conditional simulation using data and variograms for Coast Province yielded realizations with mean cumulative STH prevalences ranging from 28-42% in Coast province, and from 12-65 % in both Western and Nyanza provinces. The wider range of prevalence values simulated for Western and Nyanza province reflects the higher degree of uncertainty due to the lack of survey

points in this area. Using data and variograms for Uganda, simulations yielded cumulative STH prevalence ranging from 27-90% and 20-90% for Western and Nyanza provinces, respectively.

The trade-off between the number of schools surveyed and the proportion of districts correctly classified in each province, averaged over all sample sizes at each school, is presented in Figure 3.4a. For all provinces, and over both scenarios of spatial heterogeneity, there is a marked initial increase in the proportion of districts correctly classified with increased sampling effort. However, with the addition of extra schools there is diminishing benefit in terms of correct classification so that sampling more than four schools yields little extra performance.
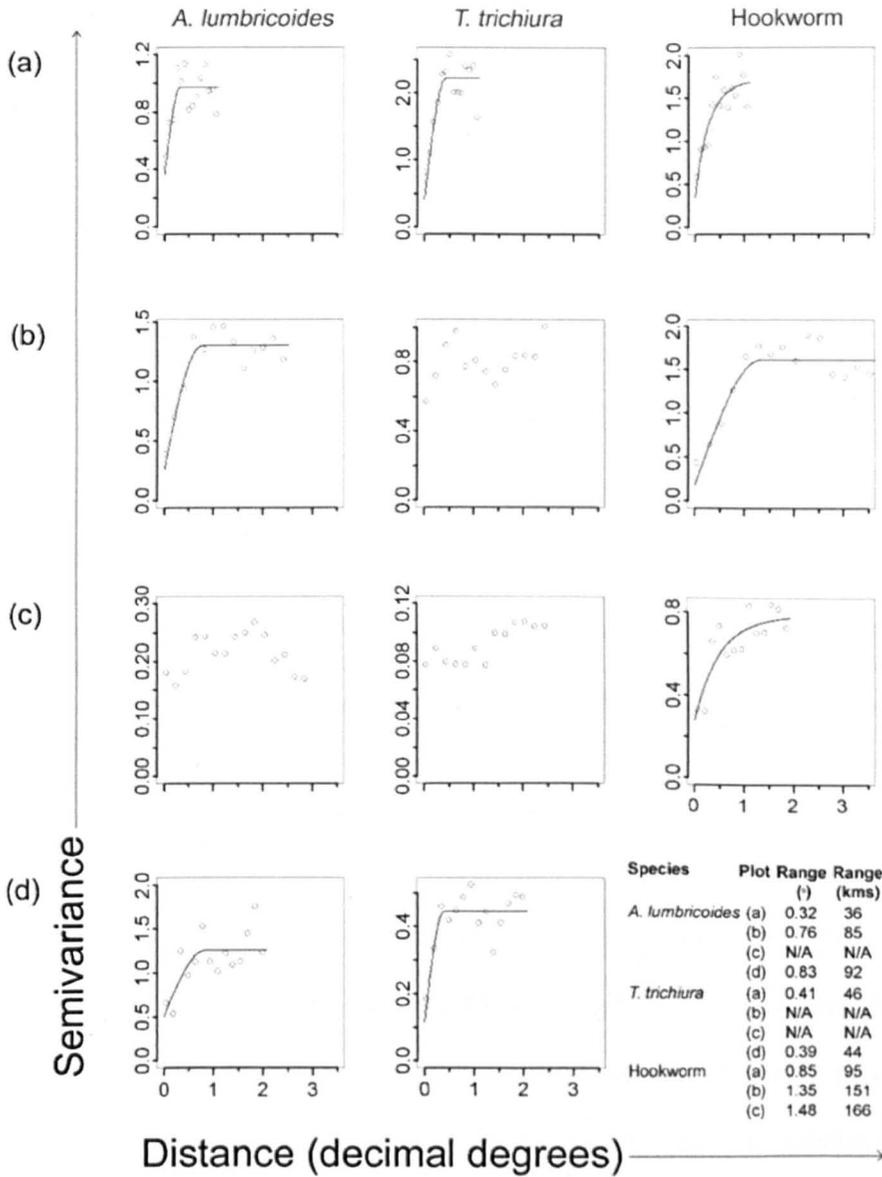
Figure 3.3 Semi-variograms and best-fitted lines of spatial models for STH: *Ascaris lumbricoides* and *Trichuris trichiura* in (a) Kenya, (b) Uganda & N. Tanzania >2° south, (c) Uganda & N. Tanzania <2° south, (d) Zambia; and hookworm in (a) Kenya, (b) Uganda & Tanzania, (c) Zambia. Vertical axis presents the semivariance and horizontal axis presented distance in decimal degrees. Range in kilometers was calculated assuming 1 decimal degree is equal to 111 km at the equator.

(a)　　　　　　　　　　　　(b)
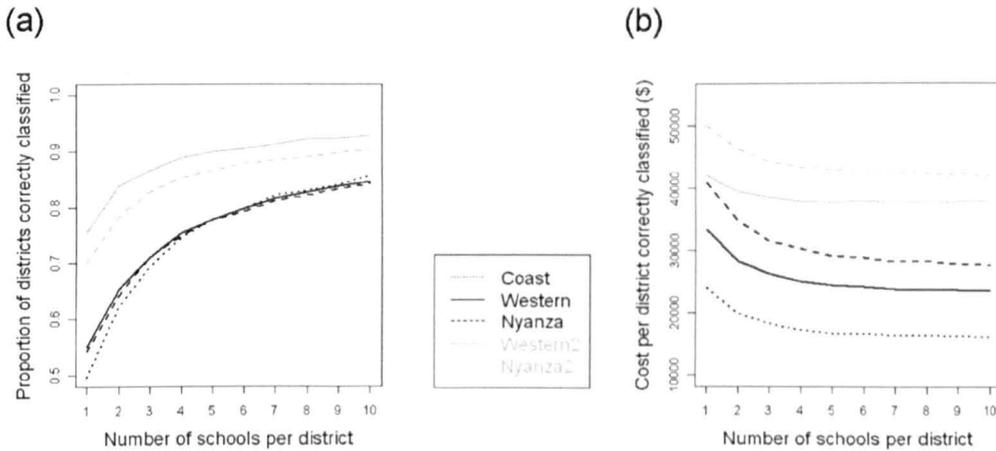
Figure 3.4 (a) Relationship between the number of schools surveyed and the ability to correctly classify districts according to treatment strategy in Coast, Nyanza and Western provinces, based on conditionally simulated data from Kenya data and variograms (Coast, Western and Nyanza) and Uganda data and variograms (Western 2 and Nyanza 2). For presentational reasons, results are averaged over the different numbers of children sampled per school since there was little effect of sample size. (b) Cost effectiveness of different sampling strategies, averaged across different number of children per school, using a treatment cost of $0.15 per person and considering 1 year of treatment.

Table 3.3 shows the range in performance using 10 and 60 children per school and 4 schools per district. Altering the number of children sampled in each school made little difference in overall accuracy. This slightly counter-intuitive result is due to the fact that accuracy is determined by the ability to classify districts according to prevalence classes, as opposed to the precision of the prevalence estimate itself. Sensitivity analysis found that the prevalence of co-infection in the 537 schools (Table 3.1) was slightly higher than would be expected by chance: hookworm and *Ascaris* (HA) co-infection being 1.1 times higher than expected by chance, hookworm and *Trichuris* (HT) 1.13, *Ascaris* and *Trichuris* (AT) 1.3 and co-infection with all three species (HAT) 1.7. Use of these probabilities led to lower estimates of cumulative STH prevalence at each school including the simulations. However, these different estimates made very little difference in performance of sampling strategies (Table 3.3).

99

Table 3.3 Performance of sampling schemes for three provinces in Kenya surveying four schools per district, based on varying assumptions of the probability of co-infection. Independent co-infection assumes co-infection prevalence equal to that which would be expected by chance and non-independent co-infection uses co-infection probability derived from analysis of the four datasets presented in Table 3.1. Grossly misclassified refers to low prevalence districts classified as high prevalence or high prevalence districts classified as low prevalence.

Province	Assuming independent co-infection				Assuming non-independent confection			
	Proportion of districts correctly classified (%)		Proportion of districts grossly misclassified (%)		Proportion of districts correctly classified (%)		Proportion of districts grossly misclassified (%)	
	10 children /school	60 children /school	10 children /school	60 children /school	10 children /school	60 children /school	10 children /school	60 children /school
Coast	71.6	76.9	0.5	0	70.3	76.1	0.2	0
Western	73.4	77.4	0.3	0	72.5	78.2	0.2	0
Nyanza	71.7	76.7	0.2	0	72.4	77.6	0.1	0
Western 2	87.0	90.1	0.02	0	85.4	89.5	0.01	0
Nyanza 2	83.1	87.0	0.01	0	81.9	86.1	0.01	0

The survey cost per school varied with the number of children sampled per school and the number of schools that could be surveyed using the same fixed costs. In Coast province, the cost to survey one school ranged from $192 when one school and ten children per school were surveyed per district, to $302 when ten schools and sixty children per school were surveyed per district. In Western province the survey cost per school ranged from $191 - $295 and in Nyanza province from $189 - $277. Figure 3.4b presents the relationship between the number of schools surveyed (averaged over the different numbers of children sampled at each school) and the total (survey

plus treatment) costs per district correctly classified assuming a treatment cost of $0.15 per individual per round and one year of treatment. For all scenarios, there is a non linear decrease in cost per district correctly classified with increasing number of schools surveyed per district. An initial increase in the number of schools surveyed led to large cost savings, whereas surveying more than four to five schools resulted in little additional improvement in cost-effectiveness. Varying treatment cost and delivery time period yielded similar conclusions (Figure 3.5). As found with performance, increasing the number of children surveyed per school made little difference to cost-effectiveness: for example, when sampling four schools per district in Coast province the (survey and treatment) cost per district correctly classified decreased from US$18,182 to US$16,579 when increasing the number of children sampled per school from ten to sixty.
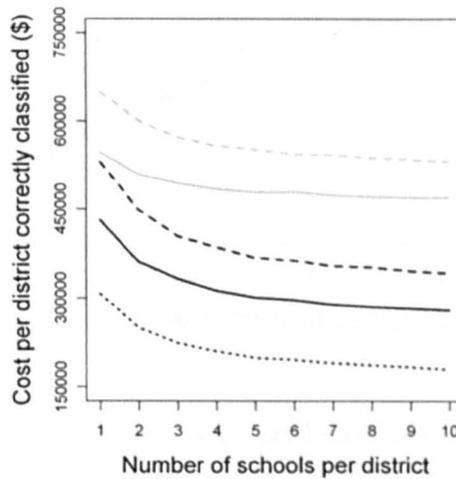
Figure 3.5 Cost effectiveness of different sampling strategies, averaged across different number of children per school, using a treatment cost of $0.39 per person and considering 5 years of treatment. Black dotted line refers to Coast province, black solid line to Western and black dashed line to Nyanza. Grey solid and dashed lines refer to results using conditionally simulated data from Uganda data and variograms in Western and Nyanza provinces respectively.

Figure 3.6 shows the survey costs and the cost of misclassification (cost of unnecessary treatment) for each sampling strategy, using Coast province and treatment costs of $0.15 over one year as an example. Whilst survey costs increase linearly, the cost of misclassification decreases non-linearly reflecting the non-linear increase in accuracy associated with increasing sample sizes.
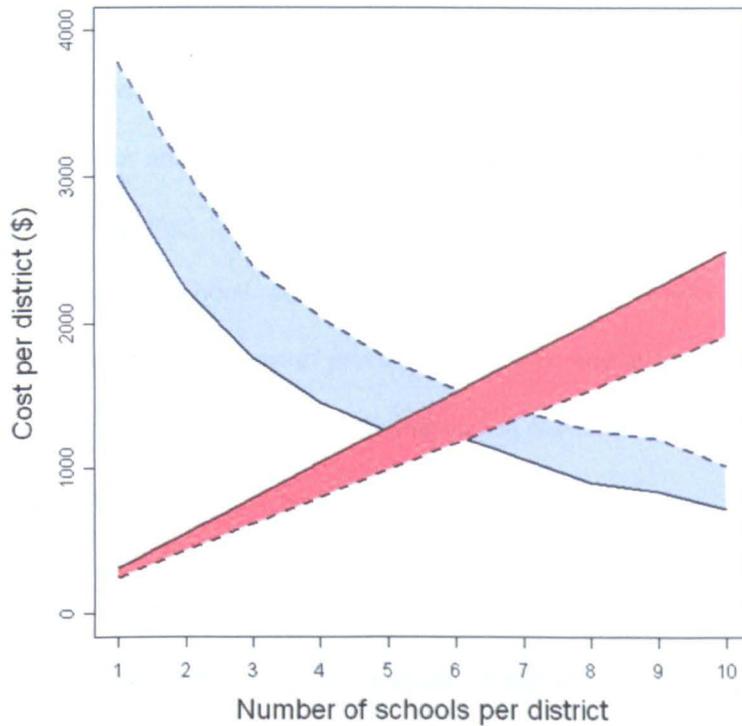
Figure 3.6 The survey cost per district of the different sampling strategies (red) and the cost of unnecessary treatment in those districts incorrectly classified in a higher endemicity class and requiring mass treatment, when mass treatment was not required (blue) in Coast Province, assuming treatment costs of $0.15 over 1 year. The solid lines represent 60 children per school and the dashed lines represent 10 children per school.

3.5 DISCUSSION

Central to the implementation of cost-effective helminth control is the need to target mass treatment to areas of greatest need. To reduce programme costs, surveys to guide such targeting should be reliable but also rapid and low cost (Brooker *et al.*, 2009a). This study represents a first attempt to account for spatial heterogeneities of STH infection when optimising sampling strategies for identifying areas requiring mass treatment. Results show that hookworm is more geographically widespread than either *A. lumbricoides* or *T. trichiura* and that for all parasites, in the datasets available for analysis, the scale of spatial autocorrelation is generally similar across different transmission settings. Using the case study of Kenya, simulation studies demonstrated that sampling four or five schools per district provides a robust method to classify districts according to prevalence across a range of prevalence scenarios and districts. Sampling more than five schools per district led to increases in performance and cost-effectiveness that are likely to be programmatically unimportant.

The results of the geostatistical analyses corroborate an earlier study in Uganda (Brooker *et al.*, 2004b) which found that spatial autocorrelation in hookworm occurred over larger spatial scales than *A. lumbricoides* and *T. trichiura*, with the latter showing small-scale or no spatial autocorrelation. The large-scale autocorrelation observed for hookworm suggest that spatially structured variables other than those included in the regression model affect hookworm transmission; possibly soil type (Mabaso *et al.*, 2003; Saathoff *et al.*, 2005). The finding that *A. lumbricoides* and *T. trichiura* show little or no autocorrelation highlights the role of small-scale, spatially stochastic variables such as differences in personal hygiene and water and sanitation. Owing to the more widespread distribution of hookworm (Brooker *et al.*, 2006b), STH infections

103

collectively are less focal than either schistosomiasis or filariasis which show autocorrelation up to

distances of <50 kms (Srividya *et al.*, 2002; Alexander *et al.*, 2003; Brooker, 2007). The

requirement of an intermediate host for schistosomiasis and a vector for LF adds a complexity to

the distribution of these parasites, necessitating the spatial congruence of human host, parasite,

and intermediate host or vector. In contrast, STHs have direct life cycles permitting transmission

where environmental conditions suit free-living parasite stages (Shope, 1999).

Such inherent differences in the spatial heterogeneity have important implications for the design

of integrated surveys that simultaneously survey STH, schistosomiasis and filariasis. The more

widespread distribution of STH implies that STH surveys can readily be integrated into surveys for

schistosomiasis and filariasis since the spatial sampling method developed for these two diseases

should sufficiently capture the spatial heterogeneities of STH infection. Current recommendations

for LF suggest that a maximum of two sites per district should be surveyed in order to assess

whether prevalence is >1%, though there is debate as to whether this sampling strategy, and the

50km x 50km grid based RAGFIL (Gyapong and Remme, 2001) are of sufficiently fine scale spatial

resolution to capture foci of infection (Srividya *et al.*, 2002; Alexander *et al.*, 2003).

Studies investigating the costs of surveys in the developing world are surprisingly sparse. A

number of studies have evaluated the cost of screening individuals for helminth infections

(Carabin *et al.*, 2000; Ansell and Guyatt, 2002) and screening versus presumptive anthelmintic

treatment (Muennig *et al.*, 1999; Brooker *et al.*, 2005). For tuberculosis, Williams *et al.* (2008)

investigated the trade-off between sampling effort and survey cost for clustered survey designs in

Cambodia. This study showed that for a given level of precision, there is a concave relationship

between cost and the number of clusters sampled so that initial increases in the number of clusters leads to a decrease in survey costs, reaching a minimum cost at 34 clusters and then rising as more clusters are sampled (Williams *et al.*, 2008). The present study is, to our knowledge, the first to evaluate the cost implications of different sampling strategies to guide treatment of STHs, whilst incorporating the cost of misclassification.

There are a number of practical implications arising from the current results. First, it is recommended that surveying four to five schools per district provides an optimal and cost-effective sampling method to guide STH control in eastern Africa. Although analysis suggests that surveying up to ten schools per district has the greatest cost-effectiveness, this benefit was minimal and surveying four to five schools provides a balance between operational ease and cost-effectiveness. In addition, increasing the number of children surveyed at each of these schools from ten to sixty makes very little difference to overall cost-effectiveness of sampling, so relatively small numbers of children per school provides a cost-effective strategy; of course, if the aim of the survey was to estimate prevalence at each school, the sample size would influence precision (Jovani and Tella, 2006). Finally, the relatively large distances over which spatial autocorrelation for hookworm occurs implies that sampling strategy developed for the more spatially focal schistosomiasis and LF will capture the spatial heterogeneities in hookworm, the most widespread STH species in much of Africa (Brooker *et al.*, 2006b).

Whilst providing a thorough examination of different sampling schemes, it is important to highlight some of the limitations of the current study. First, for conditional simulation, owing to the lack of empirical estimates, it is assumed that the spatial processes that occur in Western and

Nyanza provinces are equal to those found in either coastal Kenya or Uganda. Second, it is also assumed that, geographically, the probability of co-infection is independent (Booth and Bundy, 1995). Third, it is likely that the spatial heterogeneity of STH differ in equatorial West Africa where hookworm is often less prevalent than either *A. lumbricoides* and *T. trichiura* (Brooker *et al.*, 2006b). Further geostatistical analyses and an exploration of sampling designs in this region are required to better understand these issues. Fourth, it should be noted that the majority of the data used for the semi-variogram analysis were collected prior to the implementation of large-scale treatment programmes. It is nevertheless possible that small-scale treatment in specific, unknown, locations, may have altered the spatial heterogeneity of infection. That said, a study in Mali showed that a decade after the conclusion of a national schistosomiasis control programme, infection prevalence had returned to pre-intervention levels and showed similar patterns of spatial heterogeneity (Clements *et al.*, 2009); no comparable analysis has been undertaken to date for STH species.

In conclusion, an initial quantification of the spatial heterogeneity of STH over a number of settings in eastern Africa is presented, which shows that hookworm consistently exhibits spatial autocorrelation over larger distances than either *A. lumbricoides* or *T. trichiura*. It is further shown that sampling small numbers of children in four to five schools in each district provides a robust, quick and cost-effective sampling strategy to identify districts requiring mass treatment in an east African setting. Further work is required to investigate the cost-effectiveness of sampling in other regions of Africa and for other helminth infections, including schistosomiasis and LF, as well as malaria. Specifically, for schistosomiasis, it is unlikely that carrying out surveys using a community by community approach is feasible over large scales. Focusing on *S. mansoni*, the following chapter

aims to incorporate geostatistical methods to help inform spatial survey designs that allow decisions to be made at community level using relatively small sample sizes.

CHAPTER 4	PLANNING SCHISTOSOMIASIS CONTROL: INVESTIGATION OF ALTERNATIVE SAMPLING STRATEGIES FOR *SCHISTOSOMA MANSONI* TO TARGET MASS DRUG ADMINISTRATION OF PRAZIQUANTEL IN EAST AFRICA

4.1 INTRODUCTION

Following the methodological approach developed in the previous chapter, this chapter aims to investigate optimal survey strategies for intestinal schistosomiasis caused by *Schistosoma mansoni*. In contrast to STH, schistosomiasis is recognised as being a more focal disease (Brooker, 2007). Control strategies against schistosomiasis, which focus predominantly on MDA of the anthelmintic praziquantel are, therefore, most cost-effective when targeted to communities with the highest prevalence of infection and presumed greatest morbidity (Lengeler et al., 2002; Brooker et al., 2009a). The WHO currently recommends mass treatment of all school-age children once every two years in areas where prevalence exceeds 10% (WHO, 2006a). For *S. haematobium*, geographical targeting of treatment can be effectively and rapidly achieved through questionnaire-based studies generating data on the presence of blood in urine (a well-established marker of infection), administered by teachers to school children (Lengeler et al., 2002).

Questionnaires based on reported blood in stool are less reliable for intestinal schistosomiasis caused by *S. mansoni* (Lengeler *et al.*, 2002). Therefore, parasitological examination of stool samples remains the recommended diagnostic method, but this method is time-consuming and expensive (Brooker *et al.*, 2009a). LQAS is one approach to minimise the time and resources needed to conduct parasitological surveys and has been shown to be more cost-effective than presumptive MDA without prior surveys (Brooker *et al.*, 2005). However, LQAS requires that all schools in a given area are surveyed, necessitating significant technical and financial resources. As such, there remains a need to investigate whether *S. mansoni* surveys can be made more efficient by reducing the number of schools to be surveyed. A geostatistical approach to sampling, whereby prevalence at unsurveyed schools is predicted based on prevalence at a subset of survey schools, may offer an alternative solution.

In this chapter, two sampling designs for *S. mansoni* surveys were investigated that aim to identify schools requiring MDA in known endemic regions. Specifically, LQAS was compared to a geostatistical survey design which collects data on a subset of schools and uses this information to predict prevalence at unsurveyed schools. The ecological limits of parasite transmission were also evaluated in order to reduce the size of the sampling frame within which the two survey designs are implemented. Finally, cost estimates of both surveys and subsequent MDA campaigns were incorporated to estimate the cost-effectiveness of the alternative designs.

4.2 METHODS

4.2.1 Study settings

This analysis focuses on Oromia Regional State in Ethiopia and Western and Nyanza provinces in Kenya (Figure 4.1). These areas were chosen because of: (i) the widespread occurrence of *S. mansoni*; (ii) the availability of geo-referenced prevalence data on *S. mansoni*; and (iii) the existence of geo-referenced databases of all government primary schools. In Oromia, there are 5,251 government primary schools, and in Western and Nyanza provinces there are 5,695 government primary schools.

4.2.2 Simulation of a 'gold standard' data set

In order to generate a gold standard pseudo-dataset with realistic spatial and aspatial characteristics, data from across Kenya and Ethiopia, derived from the Global Atlas of Helminth Infection (GAHI) (Brooker *et al.*, 2000; Brooker *et al.*, 2009b), were used to investigate the spatial autocorrelation structure in observed infection patterns. Where multiple surveys from the same location were conducted at different times, the most recent survey results were used. To help standardize information, only surveys conducted in government primary schools using the WHO recommended Kato-Katz technique (WHO, 1991) were included. Data were available from 1990-2009 in Oromia and from 1992-2009 in Western and Nyanza provinces.
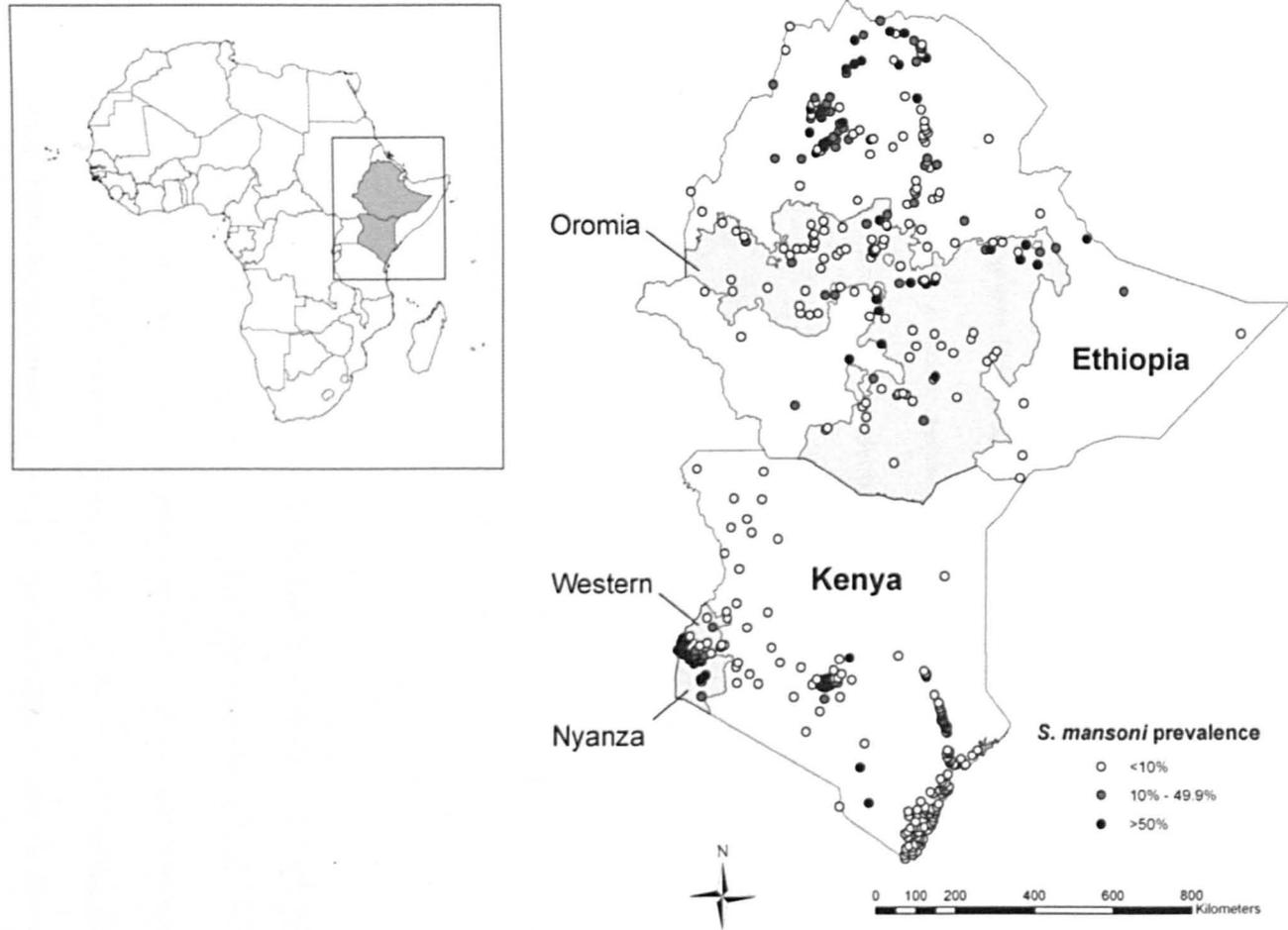
Figure 4.1 Map of surveyed primary schools where Kato-Katz was used in Kenya (n=385) and Ethiopia (n=215) included in the present analysis. Data were derived from GAHI (Brooker *et al.*, 2000, 2009). The shaded regions in each country indicate the provinces considered in this study. Inset map: Positions of Ethiopia and Kenya within Africa.

Due to the skewed nature of the prevalence data, a logistic transformation was used before analysis, $y = (\log((d+0.01) / (1-(d+0.01))))$, where d is the raw prevalence data and y denotes the transformed variable that was approximately normally distributed. As in Chapter 3, spatial autocorrelation in transformed prevalence data was investigated using an empirical semi-variogram, which describes semi-variance (half the mean squared difference between pairs of observations) as a function of lag (the distance separating the observation locations) (Goovaerts, 1997). Semi-variograms for Ethiopia and Kenya were found to be similar and data were therefore pooled to produce a single semi-variogram, providing a more stable estimate of spatial autocorrelation. There was no evidence of a large scale trend in the data.

As done for STH data in chapter 3, the semi-variogram was then used to conditionally simulate 100 different, fully enumerated realisations at all 5,695 primary schools in Western and Nyanza provinces in Kenya and all 5,251 primary schools in Oromia Regional State in Ethiopia. A population of 500 children was assumed at each school.

Pilot simulations (10,000 iterations) showed that at a prevalence of 20% (the observed overall prevalence was 18.5%), varying the number of children from 200–1000 per school made negligible difference in the precision of the prevalence estimate at each school: assuming 200 children per school, 95% of prevalence estimates from samples of 50 individuals fell between 12%-28%; whilst assuming 1000 children/school the interval was 12%-30%. Similar results were seen assuming prevalence of 10% and 5%.

4.2.3 Defining the ecological limits of transmission

As a first step and in order to reduce the number of schools to be surveyed, climate and environmental determinants of parasite transmission and intermediate snail host development and survival were identified. To capture the influence of environmental factors, data from the National Oceanographic and Atmospheric Administration's Advanced Very High Resolution Radiometer instrument were used to derive estimates of maximum land surface temperature and normalised difference vegetation index (Hay *et al.*, 2006b) for each school location. Elevation was derived from an interpolated digital elevation model from the Global Land Information System of the United States Geological Survey (Hay *et al.*, 2006b). Distance to permanent water bodies was derived in ArcMap 9.2 from an electronic map obtained from the World Wildlife Fund (Olson and Dinerstein, 2009). All available data from the GAHI were used for this task, irrespective of diagnostic technique employed, as only information on whether infection was present or absent was required. The relationships between *S. mansoni* prevalence and environmental variables were explored visually in scatter plots, which revealed distinct thresholds beyond which prevalence was <5% (Figure 4.2, Table 4.1). These thresholds were used to exclude schools from the sampling frame, reducing the number of schools from 5,695 to 4,121 in Western and Nyanza and from 5,251 to 4,448 in Oromia (Figure 4.3).
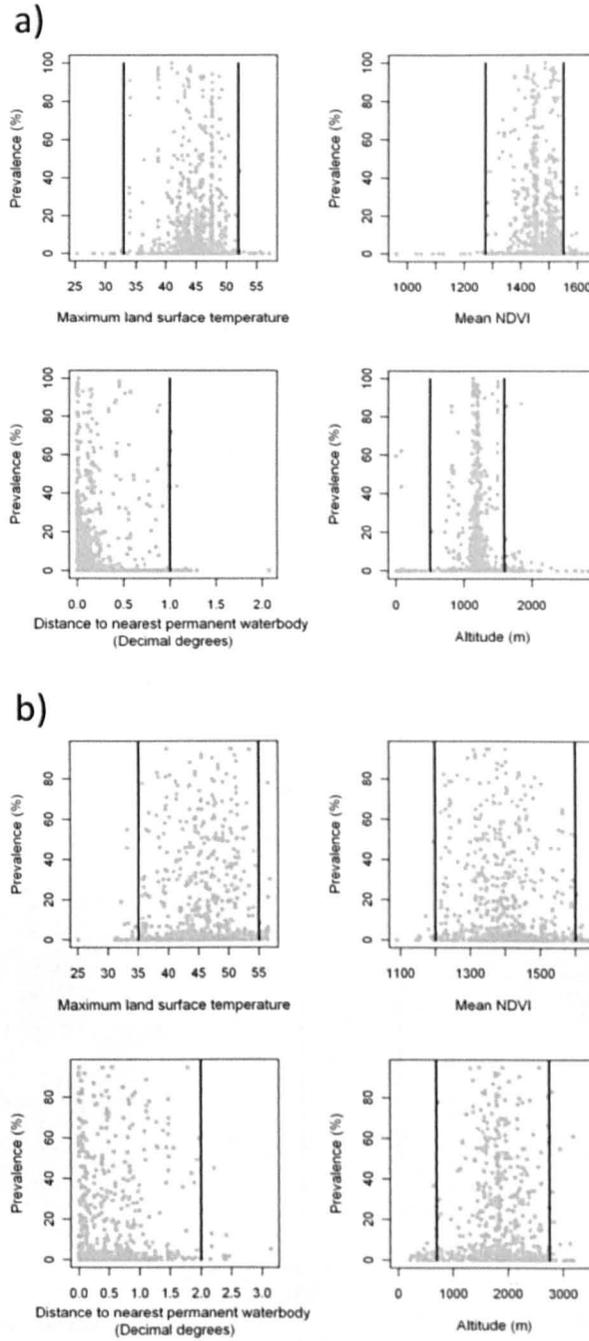
Figure 4.2 Scatterplots of the relationship between prevalence of *S. mansoni* and ecological/environmental factors, showing the limits (black lines) used define an ecological mask where prevalence is likely to be <5%, in a) Kenya and b) Ethiopia.

Table 4.1 Thresholds of the environmental variables used to describe the limits of *S. mansoni* transmission in Western Kenya and Ethiopia.

Variable	Western Kenya		Ethiopia	
	Lower limit	Upper limit	Lower limit	Upper limit
Maximum LST[1] (ºC)	33	52	35	55
NDVI (no units)	0.19	0.875	0	1
Altitude (m)	500	1600	700	2700
Distance to nearest water body (kms)	0	110	0	220

[1] LST = land surface temperature; NDVI = normalised difference vegetation index.
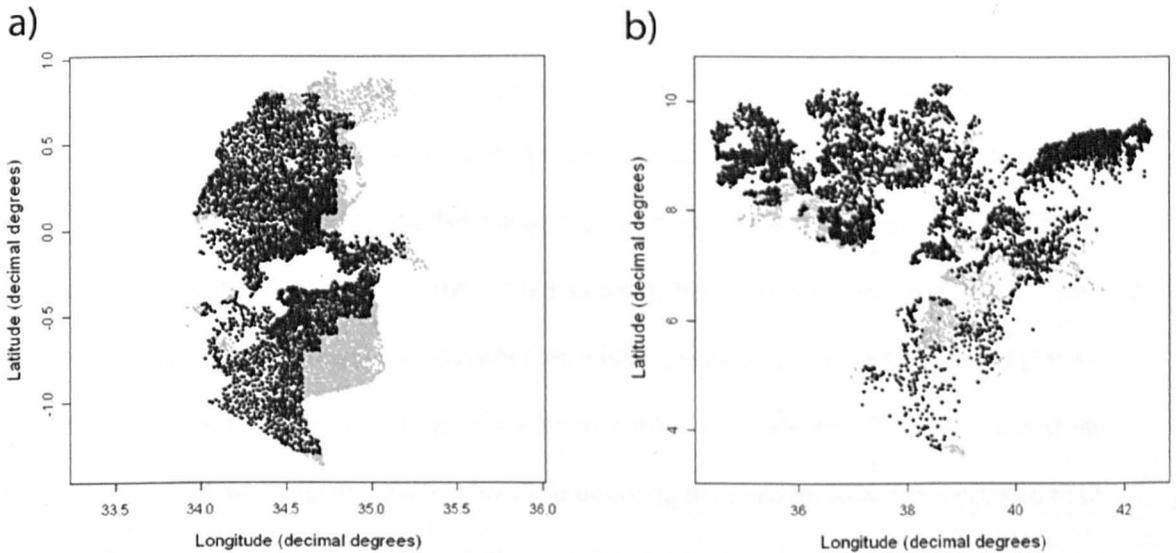
Figure 4.3 All public primary schools in a) Western and Nyanza provinces, Kenya and b) Oromia Regional State, Ethiopia. Schools shaded in grey indicate schools in areas of unlikely transmission.

4.2.4 Survey designs

Two sampling designs were considered: (i) LQAS (Lemeshow and Taber, 1991) and (ii) a variation

of the Lattice plus Close Pairs (LpCP) design (Diggle and Lophaven, 2006). LQAS required all schools

that lie within the ecological limits in each region/province to be sampled, whereas the lattice plus

close pairs design involved undertaking surveys in a sample of schools selected using a predefined

grid and using the collected empirical data to predict prevalence across all schools on the basis of

a spatial interpolation technique known as kriging (Goovaerts, 1997). Random sampling was not

considered, as this has been shown to be less efficient for spatial prediction than a regular lattice

design (Diggle and Ribeiro Jnr, 2007).

The LQAS method allows the categorization of populations based on disease prevalence using

small sample sizes for each sampling unit (Lemeshow and Taber, 1991; Robertson and Valadez,

2006). Previously, Brooker *et al.* (2005) used LQAS in Uganda to categorize schools according to *S.*

mansoni prevalence, whereby fifteen children from each school were randomly selected and if

seven or more were found to be positive, surveying was stopped and the school was classified as

having a high (≥50%) prevalence. If after fifteen samples, between two and six individuals were

found to be positive the school was classified as having prevalence ≥20 and<50%, and if fewer

than two were positive, the school was classified as having a prevalence <20%. Since this study,

WHO have revised the lower prevalence threshold denoting the need for MDA from <20% to <10%

(WHO, 2006c). Using the simulated realisations of data for Western and Nyanza and for Oromia,

this sampling plan was evaluated alongside an adapted plan using a lower stopping rule of only

one positive.

The LpCP (Diggle and Lophaven, 2006) approach surveys a subset of schools and uses the collected data and spatial interpolation methods to predict prevalence at all other unsurveyed schools. Based on a regular lattice with some additional close pairs of points, the LpCP design balances both the need to estimate semi-variogram parameters and to predict prevalence values at unsurveyed locations throughout the survey region (Figure 4.4).
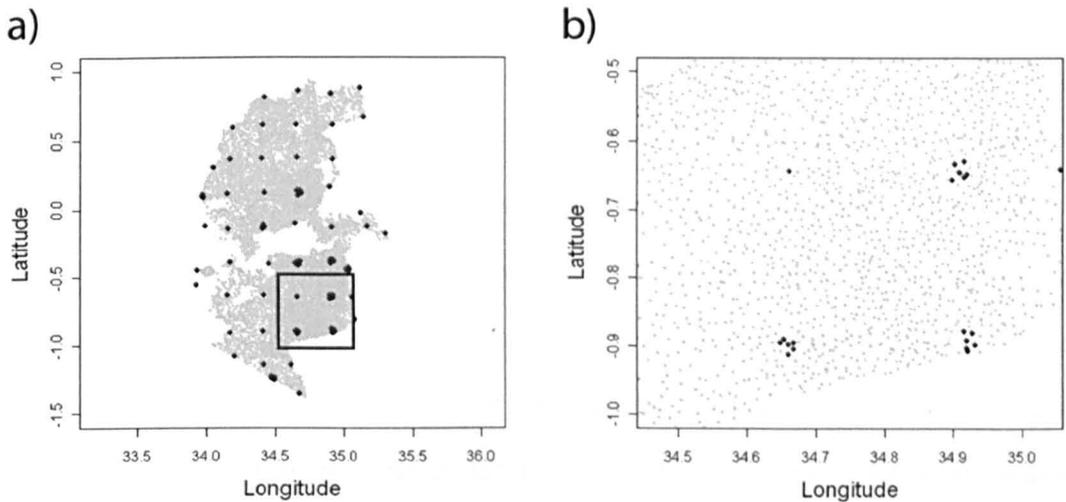
Figure 4.4 a) Illustrative example of the lattice plus close pairs design using a grid size of 27.5km in Western and Nyanza provinces, Kenya. Dark points refer to survey schools and grey points to non-surveyed schools. b) A close-up of a region (black box) showing the locations of some of the clusters of closely located schools.

The selection of sites worked as follows. A regular lattice was placed over the study region and the school closest to the nodes of the lattice was selected. For each lattice site, the distance to their five closest neighbour sites was averaged. The ten lattice sites with the shortest mean distance to their five closest neighbours were identified and the five neighbour sites were selected. This resulted in fifty additional survey sites in clusters of five, surrounding ten of the initial N lattice

sites (Figure 4.4). The inclusion of these additional sites allows for a more robust estimation of semi-variance over sites separated by very small distances, which would not be possible using a grid design alone, thus helping to infer the shape of the semi-variogram. Eight different sizes of lattice were considered: 27.5; 16.5; 13.5; 10.0; 8.0; 7.0; 5.5; and 5.0 km. Due to the large size of Oromia Regional State, lattice sizes of smaller than 10 km resulted in sample sizes of greater than 3,000 and therefore only lattices of 10-27.5 km were considered.. For each lattice size, fifty additional sites were selected as close pairs.

Once the sites had been selected using the LpCP design, the prevalence class of each school was estimated in the following steps: first, from each selected school, 50 children were randomly selected to estimate the prevalence of *S. mansoni* infection, and this estimate was used to determine whether prevalence in that school was ≥10%, thereby warranting MDA; second, prevalence values for each survey school were logistically transformed and a semi-variogram was generated, through which an exponential model semi-variogram was fitted using weighted least squares; third, the estimated semi-variogram parameters were used to predict prevalence values at all unsurveyed schools using ordinary kriging (Goovaerts, 1997). Predicted prevalence values were then used to estimate the school endemicity class (i.e. prevalence < or ≥10%).

4.2.5 Estimating survey costs

Survey cost estimates were based on actual experience of conducting field surveys in Kenya and

Ethiopia by the study authors (HJWS, RA, JHK and SB) during 2008-2009. Relevant unit costs were

identified according to an ingredients based approach (Drummond *et al.*, 2005). The quantity or

usage of each ingredient was determined and combined with cost information to produce a

monetary valuation of total resources used. Unit costs and quantities were established from the

project accounting systems in Kenya and Ethiopia and from interviews with survey staff (Table

4.2). Two categories of costs were identified: (i) imported equipment which was assumed to be

similar in both settings and excluded costs of importation; (ii) locally procured equipment, salary

and transport costs, which were incurred locally, and therefore differed between settings. Based

on our field experience, it was assumed that one supervisor, one technician and one cleaner were

required per day, irrespective of the survey design used. Consumable costs were dependent on

either the number of survey days or children sampled. Initially, it was assumed that two schools

could be visited per day when LQAS was used, because of the close proximity of schools and small

sample sizes, whereas only one school could be visited per day for the lattice design. An average

travel distance of 75km per day was assumed for both survey designs in Kenya, and 100km per day

in Ethiopia, due to the larger distance between schools in Oromia. A 10% contingency allowance

was also included in all designs. Capital costs were annuitized over the useful life of each item

using a discount rate of 3%, consistent with the recommendations of the World Bank (Walker and

Kumaranayake, 2002). Vehicle running costs only included maintenance and insurance. Costs were

estimated in local currency and their current values were converted into equivalent US$ using the

exchange rates at the time of the surveys: 70.25 Kenyan Shillings to US$1 and GBP 0.55 to US$1

(September 2008); 11.1 Ethiopian Birr to US$1 and GBP 0.68 to US$1 (May 2009)

(www.oanda.com/convert/classic). To allow comparison, all costs were converted to 2008 US$

using the US$ Consumer Price Index (http://www.bls.gov/cpi/). The effect of future inflation over the six years of the control programme was not included due to the difficulties in estimating future inflation rates in Kenya and Ethiopia.

The total cost of each sampling strategy was assumed to include the cost of the survey plus the cost of MDA over six years that would be carried out based on the survey results. Six years of treatment was considered a typical period between large scale surveys. To calculate the cost of praziquantel delivery, a recent estimate of combined delivery of praziquantel and albendazole to school-children was taken (Gabrielli *et al.*, 2006), and the reported unit cost of albendazole was subtracted, which resulted in an estimate of $0.295 per MDA round per child. Biennial MDA was considered to take place over six years. The inclusion of both survey and treatment costs takes into account the costs of misclassification arising from the alternative survey designs in terms of unnecessary treatment. The cost of presumptively treating all schools without carrying out surveys was also estimated. In addition, the total number of praziquantel treatments was estimated for each sampling strategy.

Table 4.2 Itemized cost profile of *S. mansoni* school surveys in Kenya and Ethiopia in 2008 prices (US$). Imported equipment and laboratory supplies are assumed to be constant over the two countries, whereas local supplies, salaries and transport are setting-specific.

Cost type	Unit	Unit cost Kenya (US$)	Unit cost Ethiopia (US$)
Equipment[1]	Microscopes	367.34	367.34
	Stool sieves	23.15	23.15
	Slide boxes	3.09	3.09
	Tally counters	8.62	8.62
	Jerry cans	1.44	1.07
	Pencils	0.96	0.16
	Buckets	0.96	0.83
	Wash basins	0.96	0.83
Salaries (Fixed)[2]	Supervisor	28.57	42.55
	Technician	14.29	37.23
	Cleaner	7.14	5.32
Consumables (Fixed)[3]	Disposable gloves	1.00	1.00
	Bin bags	1.14	3.06
	Liquid soap	1.42	0.63
	Paper towels	2.14	2.20
Consumables (Variable)[4]	Kato-Katz kits	0.3	0.3
	Stool pots	0.05	0.05
	Wooden spatula	0.03	0.03
	Microscope slides	0.05	0.01
	Questionnaires	0.06	0.05
	Marker pens	4.00	1.91
	Biros	2.00	1.16
Transport[5]	Transport cost per day	106.76	92.28

[1] Annuitized assuming a useful life of 4 years
[2] Fixed cost per day
[3] Fixed cost per school
[4] Variable cost, dependent on number of children
[5] Assumes an average distance of 75km per day for Kenya and 100km for Ethiopia

4.2.6 Sensitivity analysis

Sensitivity analysis was carried out to determine how sensitive overall costs are to the variation of major input parameters: (i) a higher drug delivery cost of $0.37 (Brooker *et al.*, 2008b); (ii) 20% increase in fuel prices, assuming that fuel costs account for 15% of treatment costs; (iii) increasing the number of schools visited per day when using LQAS from two to three schools; and (iv) economies of scale and 'learning-by-doing' due to scaling up the control programme resulting in a 15% reduction in costs per child treated. Additionally, simulations were run assuming a worst case scenario (higher drug cost, higher fuel cost, two schools per day using LQAS, and no economies of scale) and best case scenario (lower drug cost, fixed fuel cost, three schools per day using LQAS, and economies of scale).

4.2.7 Testing the performance of survey designs

The primary performance metric was the proportion of schools requiring treatment (i.e. those with a prevalence of ≥10% - termed intervention schools from here) correctly classified. Whilst WHO recommends a third prevalence class of ≥50%, in which treatment is given out annually, the resource requirements needed to carry out the necessary surveys and treatment mean that in reality three prevalence classes are rarely used. From a rapid survey perspective, therefore, it was deemed appropriate to use two classes of < and ≥10%. In addition to the proportion of intervention schools correctly classified, the overall proportion of schools correctly classified (with either < or ≥10% prevalence) and the proportion of infected children within intervention schools correctly classified was calculated. For the cost-effectiveness analysis, the total cost (survey plus treatment cost) per intervention school correctly classified was used for two reasons. First, by including treatment costs, it is possible to incorporate the cost of misclassifying and treating

schools that did not qualify for treatment. Second, the inclusion of non-intervention schools (prevalence <10%) could lead to misleading conclusions: for example, in a situation where 90% of schools have a prevalence of <10%, a survey design could theoretically classify no schools as requiring treatment and achieve 90% accuracy as it would have correctly classified those schools that did not qualify for treatment. Such a design would, therefore, be very cost-effective as, in addition to correctly classifying 90% of schools, it would be done at low total cost due to no treatment costs. The performance and cost-effectiveness of the alternative survey designs was evaluated against each realisation of the simulated gold standard data, and then averaged across all 100 realisations.

All the above analyses and simulations were carried out using bespoke code written in the R language 2.10 (R Development Core Team, 2008).

4.3 RESULTS

Data from a total of 600 schools from Kenya and Ethiopia were used in the analyses (Figure 4.1).

The overall prevalence was 18.5%. School level prevalence showed similar distributions in both

countries with a median prevalence in Kenya of 4.3% (range 0 – 100%) and in Ethiopia of 3%

(range 0 – 95%). Semi-variogram analysis suggested spatial autocorrelation was present up to

approximately one-third of a decimal degree (~34 km) (Figure 4.5).
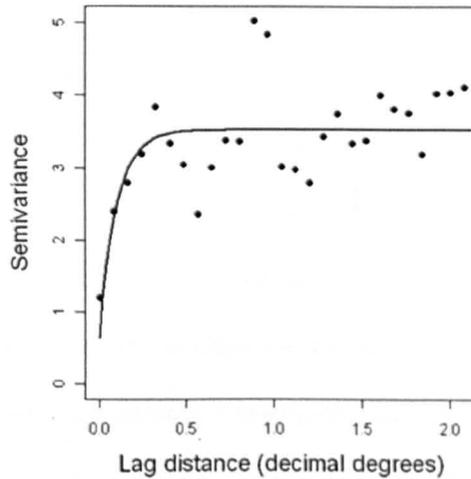
Figure 4.5 Semi-variogram of prevalence of *S. mansoni* in 600 schools across Kenya and Ethiopia. Omnidirectional semi-variogram and best-fitted line of exponential spatial model for logistically transformed prevalence data is presented. Parameter values of the fitted spatial model were range=0.31, sill=3.52, nugget=0.64. Directional semi-variograms did not differ from the omnidirectional variograms and therefore an isotropic spatial process was assumed, and an omnidirectional variogram presented. Note: at the equator, one decimal degree equates to approximately 111 kilometres.

4.3.1 Correct classification of schools

Against the derived gold standard data set, a LQAS plan using fifteen children with a stopping rule of two positives correctly classified 73.4% of schools that required intervention in Western and Nyanza and 74.3% in Oromia. A LQAS plan using a stopping rule of one positive, led to the correct classification of 88.2% of intervention schools in Western and Nyanza and 89.5% in Oromia (Figure 4.6a). On the basis of these results, a sampling plan of using fifteen children and a stopping rule of one positive was used in the subsequent comparisons with the LpCP design.

Figure 4.6a shows the performance of the different survey designs. In both settings, LQAS correctly classified a higher proportion of intervention schools than a LpCP design, with 88.4% and 89.6% correctly classified in Kenya and Ethiopia respectively. The use of smaller grid sizes in the LpCP design resulted in larger numbers of schools being selected, and consequently a higher proportion of intervention schools being correctly classified. For example, in western Kenya, reducing the grid size from 27.5km to 5km led to an increase in the number of selected schools from 91 to 776 and an increase in the proportion of intervention schools correctly classified from 51% to 73%. There was, however, a diminishing improvement in performance with increasingly small grid sizes. For a given grid size, the number of schools sampled was much larger in Oromia due to its larger size. This resulted in a larger proportion of schools being surveyed which, in turn, led to a higher proportion of intervention schools being correctly classified for a given grid size.
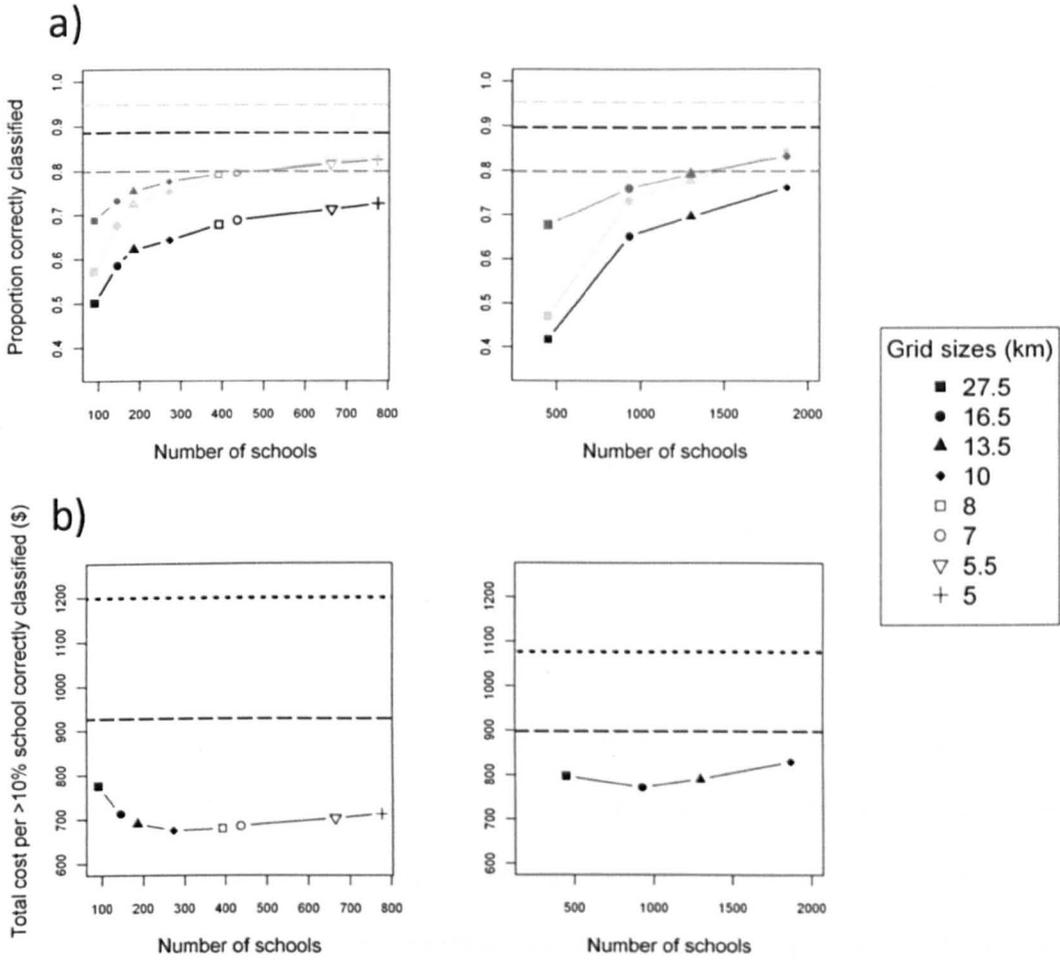
Figure 4.6 a) The proportion of intervention schools (where prevalence ≥10% and mass treatment is warranted) correctly classified using LQAS (dashed line) and a lattice plus close pairs design (black solid line) for Western and Nyanza provinces, Kenya (left) and Oromia Regional State, Ethiopia (right). Light grey lines refer to the proportion of infected children within intervention schools correctly classified using LQAS (dashed) and LpCP (solid). Dark grey lines refer to the proportion of schools (prevalence < or ≥10%) correctly classified. b) The cost-effectiveness of different survey designs using LQAS (dashed), a lattice plus close pairs design (solid) and presumptive treatment (dotted), in Western and Nyanza provinces, Kenya (left) and Oromia Regional State, Ethiopia (right). Black symbols denote the grid size used in the lattice plus close pairs design. Graphs assume 6 years of biennial treatment at a lower treatment cost of $0.295 per person. Note that the lines referring to presumptive treatment are flat as no schools are surveyed using this approach. Similarly, lines referring to LQAS are flat as all schools are surveyed using this approach.

LQAS also correctly classified a higher proportion of infected children within intervention schools than a LpCP design, with 94.6% correctly classified in Kenya and 95.2% in Ethiopia. However, in terms of any school (> or <10% prevalence) correctly classified, a LpCP design using a grid size of between 8km in Kenya and 13.5km in Ethiopia, correctly classified around the same proportion of schools as LQAS (Figure 4.6).

4.3.2 Cost-effectiveness

Table 4.3 shows the total financial costs and the total number of praziquantel doses required using the different survey designs, assuming biennial treatment over six years and the lower drug delivery cost of $0.295 per person. An estimation of the transmission limits of *S. mansoni* substantially reduced the size of the sampling frame, which was reflected in the higher cost of presumptive treatment without applying an ecological mask. Whilst presumptive treatment obviously requires no survey costs, the resource requirements in terms of praziquantel delivery are unfeasibly large in both study regions (Table 4.3). Use of either survey design resulted in lower overall cost than presumptive treatment, as praziquantel can be targeted only to schools where it is required. In terms of differences between survey designs, survey costs were considerably lower for a LpCP design than LQAS. Likewise, treatment costs were generally lower when geostatistical designs were used.

Figure 4.6b shows the cost-effectiveness of the alternative survey designs, in terms of total cost per intervention school correctly classified. An important result from these simulations is that presumptive treatment without surveys is less cost-effective than targeted treatment based on survey results. Another notable result is that the LpCP design was generally more cost-effective

than LQAS, irrespective of the number of schools that could be assessed per day using LQAS. In the

Kenyan provinces, the most cost-effective grid size appeared to be around 13.5 – 8 kms which

resulted in the selection of between 180 – 400 schools (Figure 4.6b - left). In Oromia (Figure 4.6b),

the most cost-effective LpCP design was achieved using a grid size of around 16.5 km, resulting in

the selection of around 900 schools.

Table 4.3 Comparison of resource requirements of different survey approaches, assuming a drug delivery cost of $0.295 and 6 years of biennial treatment in Western and Nyanza provinces, Kenya and Oromia Regional State, Ethiopia. For the lattice plus close pairs design, the most cost-effective grid sizes are shown.

Strategy	Total survey costs (US$)	Total treatment costs (US$)	Total costs (US$)	Praziquantel doses used
Western Kenya				
Presumptive treatment	0	2,520,480	2,520,480	8,544,000
Presumptive treatment with ecological exclusion	0	1,823,543	1,823,543	6,181,500
LQAS (2 schools per day)	422,843	820,917	1,243,760	927,590
Lattice plus close pairs (10 km grid)	57,512	610,698	668,211	696,405
Ethiopia				
Presumptive treatment	0	2,323,568	2,323,568	7,876,500
Presumptive treatment with ecological exclusion	0	1,968,240	1,968,240	6,672,000
LQAS (2 schools per day)	534,534	941,604	1,476,139	1,063,960
Lattice plus close pairs (16.5 km grid)	230,814	602,884	833,698	681,225

4.3.3 Sensitivity analysis

The results of the sensitivity analyses are shown in table 4.4. Increasing drug delivery costs, increasing fuel prices, sampling three schools per day in LQAS and the existence of economies of scale made no difference to the observation that presumptive treatment is always more expensive that LQAS or LpCP. In addition, the cost variations made little difference to the comparison between LQAS and LpCP, with the exception that in Ethiopia, increasing the number of schools surveyed per day using LQAS from two to three resulted in comparable costs for LQAS and LpCP. However, practical experience in Ethiopia suggests that surveying three schools/day would be hard to achieve due to the large distance between schools and poor road infrastructure. Surveying three schools/day is more feasible in western Kenya where schools are closer together, but this had little effect on the differences in cost estimates.

Table 4.4 Sensitivity analysis of the cost-effectiveness (US$) of alternative sampling strategies in Western and Nyanza provinces, Kenya and Oromia Regional State, Ethiopia.

Survey type	Baseline[1]	Higher drug cost	Higher fuel cost	LQAS (3 schools/ day)	Economies of scale	Best case[2]	Worst case[2]
Western Kenya							
Presumptive treatment	1,199	1,503	1,235	1,199	1,019	1,019	1548
LQAS	918	1,072	982	830	836	746	1,140
LpCP[3] (10 km)	675	831	700	675	583	583	861
LpCP (13.5 km)	690	855	714	690	593	593	884
Ethiopia							
Presumptive treatment	1,077	1,351	1,110	1,077	916	916	1,392
LQAS	907	1053	947	810	811	717	1,097
LpCP (16.5 km)	772	924	806	772	682	682	964
LpCP (13.5 km)	790	935	829	790	705	705	978

[1] Baseline costs assume cheaper drug delivery cost of $0.295 and that two schools per day can be visited using LQAS. Fuel was considered to be 15% of treatment cost and therefore an increase in the cost of fuel by 20% resulted in an increase in treatment cost of 3% (20% of 15%).

[2] A best case scenario assumes lower drug cost, fixed fuel cost, three schools per day using LQAS and economies of scale and a worst case scenario assumes higher drug cost, higher fuel cost, two schools per day using LQAS and no economies of scale.

[3] Results from the most cost-effective efficient grid sized LpCP design are shown for each country as well as 13.5km grid.

4.4 DISCUSSION

Geographically targeting the delivery of praziquantel is an essential component of schistosomiasis control. Using data from Kenya and Ethiopia, this study evaluated the cost-effectiveness of alternative survey designs for *S. mansoni*, for which no large scale rapid survey methods currently exist. The results suggest that implementing surveys to guide treatment delivery dramatically reduces both programme costs and the number of praziquantel treatments required. The results further show that while LQAS correctly classifies a greater proportion of intervention schools, the approach is more expensive and less cost-effective than a geostatistical approach, which was shown to be more cost-effective.

The decision by control programmes about how to best target MDA should be based on a consideration of available resources and desired goals of the control programmes and its targeting strategy. In the present study, cost-effectiveness is based on minimising the cost per intervention school correctly classified, but this metric may not always be the most appropriate for control programmes. For example, programmes may wish to maximise survey performance for a given amount of financial resources. Equally, a programme may wish to minimize costs to achieve a given level of performance. It should be noted, however, that no survey design will yield perfectly accurate results and therefore the decision as to which survey design to use should be based on practical and economic considerations as well as the required accuracy in classifying schools for treatment. In situations where maximising performance is more important than maximising cost-effectiveness or minimising survey time, LQAS may be favoured due to the higher proportion of intervention schools that can be correctly classified using this method. However, this comes at considerable cost, as highlighted in the current study. Future work that links computer simulations

131

to mathematical models of transmission would help to determine which survey method offers the most cost-effective strategy for the long-term control of schistosomiasis.

The design of targeting surveys should also take into account the local ecology of transmission. In some settings, for example, the prevalence of *S. mansoni* is strongly related to distance of the community/school to shoreline of large water bodies, such as, for example, certain areas of Lake Victoria (Lwambo *et al.*, 1999; Brooker *et al.*, 2001; Handzel *et al.*, 2003). Where this relationship has been established previously, this information can be used as an indication of high prevalence and help target mass praziquantel treatment. It is unlikely that a single targeting approach will be applicable to all areas and control programmes are encouraged to make effective use of local expert knowledge to augment either LQAS or the geostatistical approach. At large spatial scales, environmental data have been successfully integrated with geostatistical modeling to map the limits and broad patterns of schistosome transmission (Brooker, 2007; Simoonga *et al.*, 2009; Magalhães *et al.*, 2011). Risk mapping is, however, unable to predict the small-scale patterns of infection required for targeting control at local scales; hence the current work. A future area of research would be to integrate environmental information into survey optimization. This is an area of interest in ecological science (Hirzel and Guisan, 2002) and merits further consideration in epidemiology and public health.

.

An understanding of the spatial heterogeneity of infection was crucial to the implementation of the geostatistical design. Such designs have been previously explored for other tropical diseases: for example, the RAGFIL method for LF (WHO, 2000a). This approach recommends the selection of communities no more than 50km apart to spatially interpolate a continuous estimate of

132

prevalence over the study region (Gyapong and Remme, 2001; Gyapong *et al.*, 2002). The RAGFIL

method has been used successfully to estimate the distribution of LF in four countries in West

Africa (Gyapong *et al.*, 2002). However, in addition to some concerns that small foci of infection

may persist between interstices of a 50 × 50km grid (Srividya *et al.*, 2002), these analyses did not

incorporate estimates of survey or treatment cost, which may affect conclusions about optimal

spacing of sample locations. The importance of considering costs in survey design has previously

been investigated in the trade-off between performance and cost of different cluster survey

designs (Connelly, 2003; Williams *et al.*, 2008), the cost-effectiveness of screening versus mass

treatment (Brooker *et al.*, 2005; Gutman *et al.*, 2009) and surveys for STH (Sturrock *et al.*, 2010).

The use of simulated data, with similar spatial characteristics to that observed in the field,

provided a gold standard against which to evaluate alternative sampling designs. Without such

simulated data it would otherwise have been unfeasible to undertake the work, since empirical *S.*

mansoni data for all schools in a given region are unavailable. There are, however, a number of

study limitations worth highlighting. First, our analysis has focused on *S. mansoni*, which is the

predominant species in Ethiopia, whereas urinary schistosomiasis, caused by *S. haematobium*, is

restricted to four small foci: the lower Wabe Shebele valley, western Welega and lower and

middle Awash valley (WHO, 1987; Kloos *et al.*, 1988). In countries where *S. haematobium* is

common, WHO recommends the use of blood in urine questionnaires, often implemented through

the education system (Clements *et al.*, 2008a), as a means for identifying high prevalence schools.

What has hindered the control of intestinal schistosomiasis is a lack of rapid assessment, the issue

addressed by the present study. Information elicited from a blood in urine questionnaire survey

will need to be combined with data from rapid *S. mansoni* survey to develop an overall national

schistosomiasis control strategy, as relying on survey data from only one species can result in missing foci of infection (Gutman *et al.*, 2008).

A second limitation is that the spatial heterogeneity of *S. mansoni* infection may differ in other regions, making the extrapolation of conclusions to other settings difficult. Encouragingly, however, previous spatial analysis of *S. mansoni* in Cameroon, Mali and Uganda showed remarkably similar scales of spatial heterogeneity as observed in the present study, with clustering occurring up to around 50km (Brooker, 2007). Equally, semi-variograms estimated from survey data collected from Rwanda and Tanzania, extracted from GAHI, indicates that *S. mansoni* appears to cluster at distances of 40-132kms (Figure 4.7). Such consistency in clustering suggests that grid sizes of between 10-16.5km would sufficiently capture the spatial heterogeneity of infection across sub-Saharan Africa.
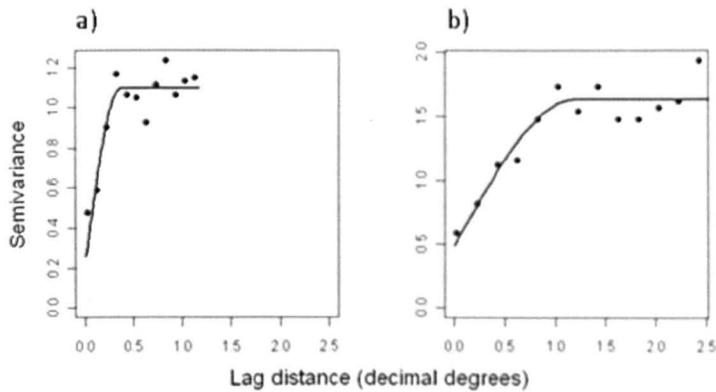
Figure 4.7 Semi-variograms of prevalence of Schistosoma mansoni in a) 133 schools in Rwanda and b) 143 schools in Tanzania. Detrended omnidirectional semi-variograms and best-fitted lines of spherical spatial models for logistically transformed prevalence data are presented. Note: at the equator, one decimal degree equates to approximately 111 kilometres.

Third, it was not possible to explore the implications of prevalence on survey design. In situations where prevalence is very high, targeting treatment at the school level via surveys is likely to become less cost effective, as most schools will qualify for treatment. In such circumstances, conducting surveys at the sub-district/district level or even using presumptive treatment without surveys may be more cost-effective (Brooker *et al.*, 2005; Gutman *et al.*, 2009). Repeat analyses in different epidemiological settings would be useful.

A fourth limitation is that it is possible that the spatial characteristics of infection may also vary over time, due to changes in ecology, demography and introduction of MDA campaigns, which could affect the performance of any geostatistical design over time. That said, recent work comparing *S. mansoni* infection in Mali suggested that 12 years after a 10 year national drug

campaign, the spatial distribution of infection was similar to that seen pre intervention (Clements *et al.*, 2009). Fifth, it should be noted that implementation of the LpCP design requires knowledge of the locations of schools in order to aid the selection process and prediction stages. Encouragingly, however, an increasing number, perhaps even a majority, of ministries of education in Africa have georeferenced school databases as part of their Education Information Management Systems.

Results show that the optimal grid size varies according to the spatial density of schools, such that grid sizes should be chosen appropriate to the study area: where schools are sparsely distributed, as in Oromia Regional State, a larger grid size (16.5km) is more cost-effective; whereas, where schools are more densely distributed, a finer grid (up to 10km) seems more appropriate. As a compromise, a grid size of 13.5km could be used in the current study areas. As the performance of such a geostatistical design is likely to vary between settings due to differences in infection prevalence, ecology, and distribution of the population, this approach warrants further investigation and validation in the field. A potential drawback of our geostatistical approach to targeting praziquantel is the technical requirements to implement the initial modelling to parameterise the sampling design. Many national schistosomiasis control programmes lack epidemiologists and this hinders several aspects of programme implementation: for example, the design of rigorous monitoring and evaluation strategies. Indeed, large-scale implementation of LQAS should be preceded by some form of validation of sampling schemes, often undertaken using a combination of computer simulation and field studies. To overcome the lack of technical capacity, national programmes often draw upon regional and international expertise. No one would dispute national programmes asking for a health economist to design an economic

evaluation; in the future, programmes may request technical assistance in mapping and geostatistical modelling. Furthermore, national programmes may wish to develop their own such capacity and with the growing availability of open-access spatial tools, such as GRASS GIS and R statistical package, this will hopefully become an increasingly viable option. Importantly, adding the cost of any external technical assistance (~US$10,000) would not change the overall conclusions of the study.

In summary, using a computerized simulation approach, it is shown that targeting praziquantel at school/community level is more cost-effective than presumptive treatment for the control of *S. mansoni* in East Africa. It is further shown that while LQAS correctly classifies a greater proportion of intervention schools, a geostatistical approach is more cost-effective. Control programmes should consider the trade-offs between maximizing the numbers of infected individuals who receive treatment and how best to use their limited resources – an inevitable feature of public health programmes.

CHAPTER 5 | WHAT SPATIAL SCALE SHOULD BE USED IN THE INTEGRATED MAPPING OF HELMINTH AND *PLASMODIUM FALCIPARUM* INFECTIONS?

5.1 INTRODUCTION

The previous chapters investigated alternative survey approaches for STHs and *S. mansoni* to identify where MDA is required by determining whether prevalence of infection exceeds some designated threshold (Sturrock *et al.*, 2009; Sturrock *et al.*, 2010). In these analyses, the geographical level used for mapping was assumed to reflect the IU at which the decision to deliver mass treatment is taken. For example, for STH the IU is normally the district (second administrative unit) and for schistosomiasis the school or community. It is unclear, however, whether these units offer the most suitable spatial scales for the implementation of control activities. Choice of IU, and hence geographical level of mapping, will depend on the distribution of infection: if infection is highly focal and shows considerable heterogeneity over small areas, a lower administrative unit should be chosen for the IU; whereas if the infection is more widespread, a larger administrative unit should be chosen.

Despite the importance of spatial scale for disease mapping, and the implementation of control, work on this topic is noticeably sparse. In Nigeria, King *et al.* (2009) assessed whether school-level mapping of schistosomiasis could also be used for implementing trachoma interventions and conversely, whether district-level mapping of trachoma could be used for targeting praziquantel treatment for schistosomiasis. The authors concluded that whilst both integrated designs were easy to implement, school surveys were more informative in determining intervention strategies than district-level methods, as a greater number of communities that warranted treatment for both trachoma and schistosomiasis were missed using district-level mapping compared to school-level mapping. For LF, the IU is generally the second administrative level – the district (WHO, 2005). This level was identified, in part, through spatial analysis by Gyapong and Remme (2001) who investigated a rapid mapping method based on a spatial sampling grid with 50km between sampled villages showing that such a design yielded operationally similar prevalence categories to those obtained with a 25 x 25km grid. Other work in India and Papua New Guinea has, however, suggested that foci of filarial infection may occur between the interstices grid these sizes (Srividya *et al.*, 2002; Alexander *et al.*, 2003).

A further factor which may influence the geographic level which should be used for mapping is the underlying prevalence of infection. Where prevalence is very high, a higher proportion of communities/schools will require MDA, irrespective of the scales over which disease cluster. In such settings, conducting mapping and control activities over large spatial scales, or even treating entire populations presumptively, may provide a more cost-effective approach (Brooker *et al.*, 2005; Gutman *et al.*, 2009). In contrast, in settings where infection is less prevalent, there will be fewer foci of infection and mapping at a higher spatial resolution would be required to identify

hotspots and implement control. For example, where malaria transmission is low and concentrated around focal points, such as in the highlands of Kenya, targeted implementation of IRS provides a more feasible and effective alternative to mass spraying of entire areas (Zhou *et al.*, 2010).

This chapter investigates the implications of implementing mapping surveys and subsequent interventions at varying spatial scales for *S. mansoni*, STH infection and *P. falciparum*. As well as identifying the optimal spatial scale over which surveys for each species should be conducted, this study aims to assess whether there is a single scale at which surveys for each of these species can be validly integrated. The effect of infection prevalence on survey design is also explored. Mapping of malaria is included because schools are increasingly being used in surveys (Brooker *et al.*, 2009c; Gitonga *et al.*, 2010; Ashton *et al.*, 2011) and there is growing interest to implement integrated helminth and malaria control in schools (Brooker *et al.*, 2006c; Brooker *et al.*, 2007). Analysis follows that of previous chapters, employing spatial analysis, conditional simulation and economic evaluation.

5.2 METHODS

5.2.1 Study area

The current analysis was undertaken among all government primary schools in four provinces in Kenya: Nyanza, Western, Coast and Eastern provinces (Table 5.1). These provinces were chosen due to the availability of a georeferenced database of public primary schools, varying levels of endemicity and varying sizes of districts and sub-districts.

Table 5.1 Summary of provinces used in the simulations.

Province	Number of schools	Number of districts[1]	District area (range kms^2)	Number of sub-districts	Sub district area (range kms^2)
Nyanza	3,728	12	579 – 2,109	33	32 – 1,005
Western	2,046	8	548 – 2,072	24	94 – 699
Coast	1,095	7	246 – 39,161	22	12 – 15,270
Eastern	4,233	13	730 – 61,126	38	197 – 37,912

[1]As of 1999

5.2.2 Generation of gold standard data

In order to generate as realistic prevalence data as possible for each infection, it was first necessary to gain an understanding of the spatial and aspatial variance characteristics of the different diseases across different epidemiological settings. To do this, nationwide survey data from Kenya were used: data on STH (*Ascaris lumbricoides*, *Trichuris trichiura* and hookworm) and

141

S. mansoni infection were extracted from the GAHI (www.thiswormyworld.org); data on *P. falciparum* infection were obtained by a 2008-2009 national school survey (Gitonga *et al.*, 2010) (Table 5.2). To help standardize information on helminth species only school surveys where Kato-Katz was the diagnostic method used were included. As done in previous chapters, large scale trends were removed by including longitude and latitude in a logistic regression model. To account for clustering of infection at the school level, school was included in the model as a random effect. Semi-variograms were then estimated using the school level random effect at each location.

Table 5.2 Summary of school prevalence survey data used for spatial analysis

Parasite	Number of sites	Mean prevalence (%)	School level range in prevalence (%)
S. mansoni	385	13.5	0 - 100
Hookworm	425	23.4	0 – 95
A. lumbricoides	425	14.3	0 - 88.6
T. trichiura	425	20.5	0 - 98.1
P. falciparum	529	7.6	0 – 88.5

As in previous chapters, 1000 realisations of fully enumerated gold standard data were generated using the semi-variogram parameters and conditional simulation. Where any large scale trends were present, these were added to the conditionally simulated data. As in chapter 3, cumulative prevalence of STH infection (*p*, prevalence of any one of the three STH species) at each school was

estimated assuming independence of co-infection as follows: $p = H + A + T - (HA) - (HT) - (AT) + (HAT)$ (Booth and Bundy, 1995), where H was the proportion infected with hookworm, A the proportion infected with *A. lumbricoides* and T the proportion infected with *T. trichiura*.

5.2.3 Intervention thresholds and strategies

For *S. mansoni*, biennial MDA of praziquantel is recommended in schools where prevalence is ≥10% and for STH, annual MDA of albendazole should take place in schools where prevalence of any STH species is ≥20% (WHO, 2006c). There is no similar prevalence threshold above which malaria interventions are recommended to be implemented, therefore, the recommendations by Hay *et al.* (2008) were used, who suggest that there should be total coverage of available interventions in areas where prevalence is >5%. For the purpose of this study, it was assumed that where survey results observe a prevalence exceeding 5%, 1 year of four-monthly school-based IPT (Clarke *et al.*, 2008; Temperley *et al.*, 2008) would be implemented.

5.2.4 Survey approaches

The following survey approaches were compared:

Lot Quality Assurance Sampling (LQAS)

As used in chapter 4, this method uses a small representative sample of a study population to identify whether the number of infected individuals (prevalence of infection) is above or below a designated threshold (Lemeshow and Taber, 1991). LQAS decision rules are based on the cumulative binomial probabilities that a designated number of individuals (d) are found to be

infected from a given number of individuals (n) in a sample. If d is exceeded before n is reached, then the community is classified as high risk and sampling can stop; if n is reached before d is exceed then the community is defined as having prevalence below the designated threshold. Values for d and n will depend on both the prevalence threshold of interest and the level of error deemed acceptable by the user.

For *S. mansoni*, the results presented in chapter 4 showed that a sampling plan of $n=15$, $d=1$ provides a robust method to identify schools that have a prevalence of ≥10% and therefore require treatment. Other previous work in Uganda showed that a plan of $n=15$, $d=2$ provided good levels of performance in identifying schools with a prevalence of *S. mansoni* ≥20% (Brooker *et al.*, 2005). As recommendations for MDA for STH also use 20% as a threshold, this same sampling plan was evaluated for cumulative STH here.

For *P. falciparum*, LQAS has previously been used in Madagascar to identify schools with a prevalence of ≥15% (Rabarijaona *et al.*, 2001). The authors showed that using a sample size of 36 and a threshold of 2 positives, provided a sensitivity of 100% and specificity of 86%, when compared to a gold standard sample of 70 children. Here, LQAS is used to identify schools where prevalence is >5%, the prevalence threshold suggested by Hay *et al.* (2008). Four different LQAS sampling plans were evaluated: ($n=35$, $d=1$), ($n=30$, $d=1$), ($n=25$, $d=1$) and ($n=20$, $d=1$).

Sub-district approach

This approach aimed to allow decisions on intervention to be taken at the sub-district level (administrative level 3). Using this method, n schools were selected in each sub-district, with 50 children randomly selected in each school. This number of children in each school was chosen as it is the current recommended sample size to be taken at school level during surveys for schistosomiasis and STH (WHO, 2006c). Prevalence of infection at the sub-district level was then calculated. If this prevalence exceeds the designated threshold for a given disease, then all schools in the sub-district would receive the relevant intervention. This method results in less sampling effort than LQAS (which requires all schools to be sampled), but has a higher chance of missing foci of infection or mistreating large numbers of school children in schools not requiring intervention. For all species the performance of randomly selecting 1, 2, 3, 4, 5 and 10 schools per sub-district was tested. These were thought to be an operationally feasible range of sample sizes.

District approach

The district approach was identical to the sub-district approach, however, the process is carried out at the district level (administrative level 2). This method requires less sampling effort than a sub-district approach, but has an even higher chance of missing foci of infection or mistreating large numbers of school children in schools not requiring intervention. To allow comparison with a sub-district approach, in each district, a random selection of 1, 2, 3, 4, 5 and 10 schools was evaluated. Figure 5.1 illustrates the three different survey methods.
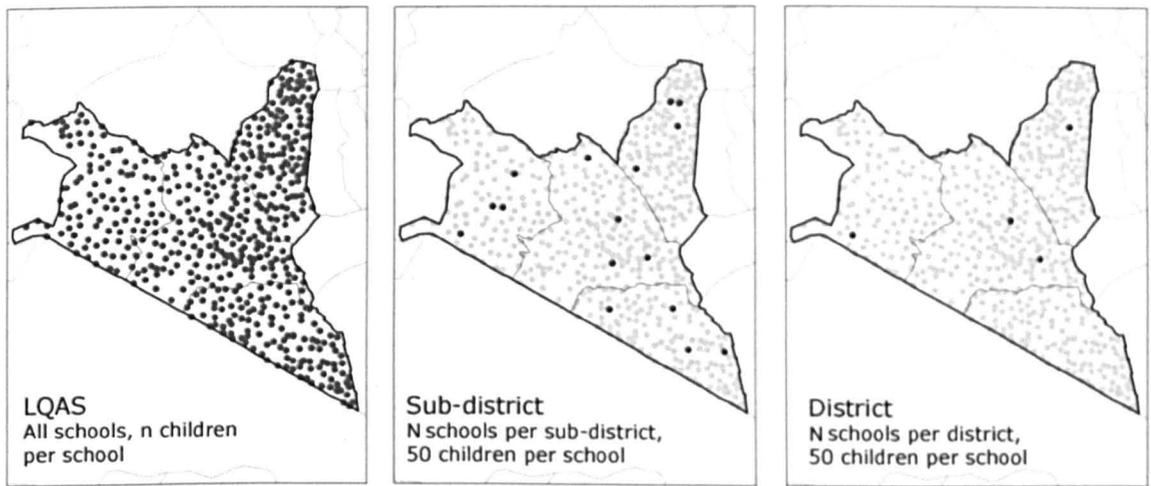
Figure 5.1 An illustration of the different survey methods investigated using an example of Migori district, Nyanza province, Kenya: LQAS in all schools, a sub-district approach using 4 schools per sub-district and a district approach using 4 schools per district.

5.2.5 Cost estimates

In a similar way to previous chapters, the cost of each survey design was estimated using itemised costs incurred during surveys carried out in coastal Kenya (Chapter 2) (table 5.3). Capital costs were annualized over the useful life of each item using a discount rate of 3% and a 10% contingency allowance was included. It was assumed that two schools could be visited per day when LQAS was used, because of the close proximity of schools and small sample sizes, whereas only one school could be visited per day when either a sub-district or district approach were used.

The total cost of each sampling strategy was assumed to include both the cost of the survey plus the cost of the relevant intervention. For STH and *S. mansoni* it was assumed that biennial MDA of albendazole and praziquantel would be carried out for 6 years in schools/areas classified as requiring treatment. Six years of MDA was considered a typical period between large scale surveys. To calculate the cost of praziquantel delivery, a recent estimate of combined delivery of praziquantel and albendazole to school-children was taken (Gabrielli *et al.*, 2006) and the reported unit cost of albendazole subtracted. This resulted in an estimate of $0.295 per MDA round per child. For STH, the cost of treatment and delivery of albendazole was calculated using a recent estimate of $0.15 per delivery round per child (Hall *et al.*, 2009). For *P. falciparum*, it was assumed that in schools/areas where intervention was thought to be required, 1 year of four-monthly IPT would take place at a financial cost of $1.20 per child treated per year (Temperley *et al.*, 2008).

5.2.6 Sampling simulations

Simulations of each of the three different survey approaches were carried out on each realisation for each disease in the four provinces. For each realisation, the performance - defined as the proportion of schools that qualify for intervention (termed intervention schools from here) correctly classified (equivalent of sensitivity) - was calculated. Additionally, the cost-effectiveness, defined as the total cost (survey plus intervention cost) per intervention school correctly classified, was calculated. Performance and cost-effectiveness results were then averaged over all realisations.

5.2.7 Sensitivity analyses

In order to test the cost-effectiveness of the different survey approaches over different cost settings, a cost sensitivity analysis was carried out using higher treatment costs and assuming that three schools per day could be visited using LQAS. For *S. mansoni*, we assumed a higher treatment cost of $0.37 per child treated per round and for STH a cost of $0.39 (Brooker *et al.*, 2008b). For *P. falciparum*, it was assumed that every child that would receive IPT would be given a long-lasting insecticidal net at a unit cost of $3.90 per net (Kolaczinski *et al.*, 2010), making the total cost of treatment $5.10 ($1.20 + $3.90).

Table 5.3 Itemized costs of conducting school surveys for STH, *S. mansoni* and *P. falciparum* infections in Kenya and Ethiopia in 2008 US$. The number of units required for 1 school of 50 children are shown.

Cost type	Unit	Unit cost (US$)	Units required for STH	Units required for S. mansoni	Units required for P. falciparum
Equipment	Microscopes	367.34	2	2	N/A
	Stool sieves	23.15	6	6	N/A
	Slide boxes	3.09	1	1	N/A
	Jerry cans	1.44	2	2	2
	Pencils	0.96	5	5	5
	Tally counters	8.62	4	4	N/A
	Buckets	0.96	2	2	2
	Wash basins	0.96	2	2	N/A
Salaries (Fixed)[1]	Supervisor	28.57	1	1	1
	Technician	14.29	1	1	1
	Cleaner	7.14	1	1	N/A
Salaries (Variable)[2]	Technician	14.29	1	1	1
Consumables (Fixed)[1]	Disposable gloves	5.00	0.2	0.2	0.2
	Bin bags	5.71	0.2	0.2	0.2
	Liquid soap	7.14	0.2	0.2	0.2
	Paper towels	2.14	1	1	1
Consumables (Variable)[3]	Kato-Katz kits	0.3	55	55	N/A
	Stool pots	0.05	55	55	N/A
	Wooden spatula	0.03	55	55	N/A
	Microscope slides	0.05	55	55	N/A
	Questionnaires	0.06	55	55	55
	Marker pens	4.00	0.2	0.2	0.2
	Biros	2.00	0.2	0.2	0.2
	Malaria RDTs	0.82	N/A	N/A	55
	Lancets	0.02	N/A	N/A	55
Transport[4]	Fuel and maintenance (per km)	1.43	75	75	75

[1] Fixed cost per day
[2] Variable cost, dependent on number of children
[3] Variable cost, dependent on number of children. N.B. 10% contingency stock added.
[4] Assumes an average distance of 75km per day.

5.3 RESULTS

5.3.1 Spatial heterogeneity and simulated data

Figure 5.2 shows the species specific semi-variograms for each parasite. Two main results are evident. Firstly, the most prevalent STH species, hookworm, shows slightly larger ranges than *A. lumbricoides* and *T. trichiura*. Secondly, *S. mansoni* appears to cluster over smaller scales than *P. falciparum*. It should be noted, however, that despite including longitude, latitude and land surface temperature both as linear and quadratic terms in logistic regression models, there is still evidence of a large scale trend in the *P. falciparum* infection data.
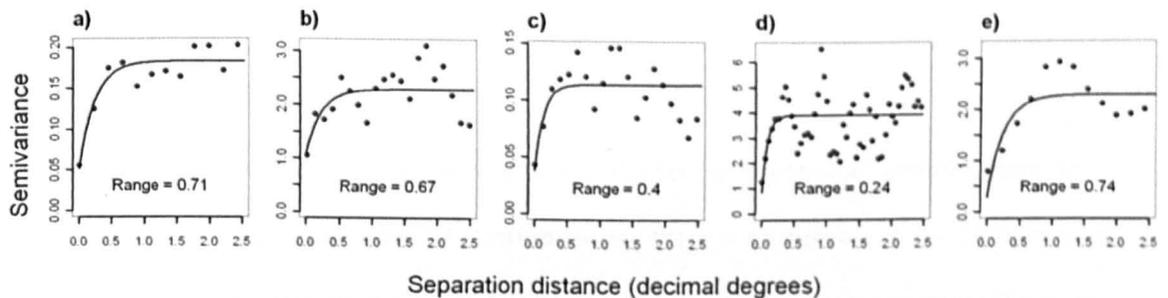
Figure 5.2 Semi-variograms for a) hookworm, b) *A. lumbricoides*, c) *T. trichiura*, d) *S. mansoni* and e) *P. falciparum* in Kenya. One decimal degree is approximately 111km at the equator.

Details of the simulated datasets for each province are shown in table 5.4. Prevalence of the three diseases was generally high in Nyanza and Western provinces, with Coast and Eastern provinces having varying levels of prevalence between parasite species.

Table 5.4 Summary of the 1000 conditionally simulated realisations for *S. mansoni*, STH and *P. falciparum* in the four study provinces.

Province	Mean *S. mansoni* prevalence (%) (range)	Mean cumulative STH prevalence (%) (range)	Mean *P. falciparum* prevalence (%) (range)
Nyanza	15.2 (6.4 – 29.5)	69.9 (60.9 – 81.4)	23.2 (19.4 – 27.6)
Western	14.9 (5.1 – 36.7)	74.1 (67.8 – 80.7)	33.4 (28.1 – 38.8)
Coast	6.6 (3.4 – 11.9)	69.2 (66.1 – 75.3)	5.0 (4.0 – 6.6)
Eastern	18.1 (13.7 – 24.3)	40.4 (35.2 – 47.6)	1.7 (0.7 – 4.4)

5.3.2 LQAS for P. falciparum

The performance and cost-effectiveness results from the four different LQAS sampling plans for *P. falciparum* are summarized in table 5.5. Unsurprisingly, there is an increase in the proportion of intervention schools correctly classified (sensitivity) as the sample size is increased. As the threshold number of positives remained one over all sampling plans, larger sample sizes also resulted in lower specificity. Sampling plans achieved higher levels of sensitivity in Nyanza and Western province than in Coast and Eastern. This is most likely due to the fact that prevalence is lower in Coast and Eastern provinces and therefore a higher proportion of intervention schools have a prevalence of very close to 5% (table 5.4), which makes them vulnerable to being classified as non-intervention schools. Cost-effectiveness appears to be maximised using the smallest sampling plan of (20,1). A sampling plan of (25,1) appears to regularly achieve good levels of performance (>88% sensitivity) across settings and is comparatively cost-effective. For this reason,

this sampling plan was used to make comparisons with a sub-district and district approach from here on.

5.3.3 Sampling simulations

Performance results from the simulations of the three survey approaches for each species are shown in Figure 5.3. For *S. mansoni* and *P. falciparum*, LQAS achieved a higher level of performance (proportion of intervention schools correctly classified) than either a sub-district or district approach for all parasites, and a sub-district approach performed slightly better than a district approach. For STH, both a sub-district and district approach, using at least three schools per sub-district/district, achieved equally as good performance as LQAS in the settings explored. For all species, increasing the number of schools surveyed per sub-district/district led to a decreasingly small improvement in performance across all species, so that surveying beyond three to four schools per sub-district/district led to relatively small improvements in performance.

Table 5.5 Performance and cost-effectiveness of different LQAS sampling plans for *P. falciparum* using simulated data at all public primary schools in four provinces in Kenya.

Province	Sampling plan	True high/LQAS high	True low/LQAS low	True high/LQAS low	True low/LQAS high	Cost per intervention school correct ($)
Nyanza	(20,1)	0.95	0.81	0.05	0.19	776
	(25,1)	0.97	0.77	0.03	0.23	787
	(30,1)	0.98	0.74	0.02	0.26	798
	(35,1)	0.99	0.71	0.01	0.29	809
Western	(20,1)	0.97	0.66	0.03	0.34	704
	(25,1)	0.98	0.60	0.02	0.40	708
	(30,1)	0.99	0.55	0.01	0.45	712
	(35,1)	0.99	0.51	0.01	0.49	716
Coast	(20,1)	0.87	0.79	0.13	0.21	1,232
	(25,1)	0.91	0.76	0.09	0.24	1,271
	(30,1)	0.94	0.72	0.06	0.28	1,314
	(35,1)	0.96	0.69	0.04	0.31	1,353
Eastern	(20,1)	0.83	0.88	0.17	0.12	2,587
	(25,1)	0.88	0.85	0.12	0.15	2,677
	(30,1)	0.92	0.83	0.08	0.17	2,790
	(35,1)	0.94	0.81	0.06	0.19	2,909
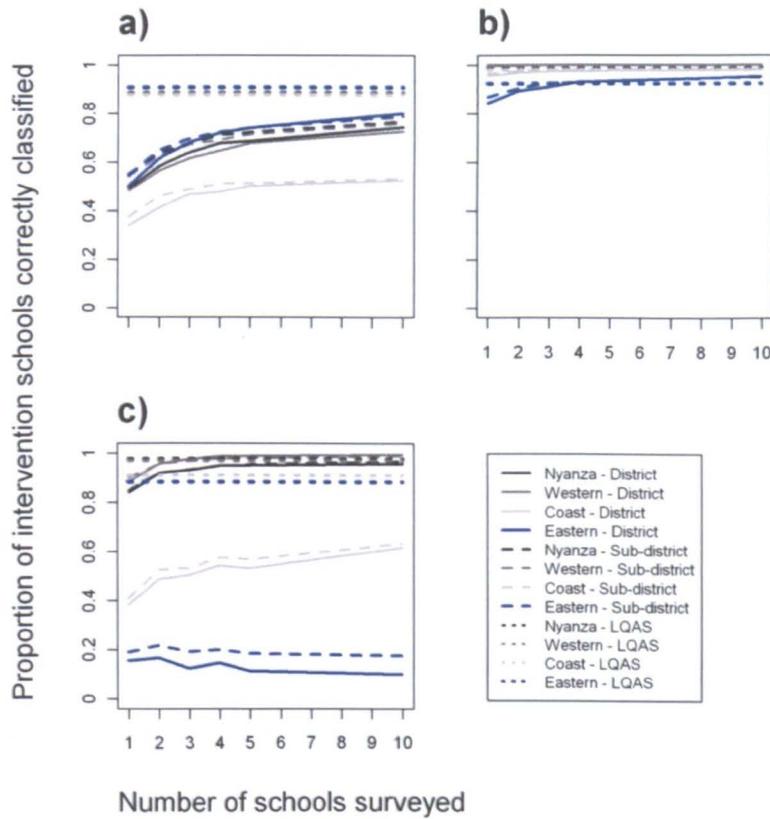
Figure 5.3 The performance of different survey approaches for a) *S. mansoni*, b) STH and c) *P. falciparum*.

Figure 5.4 compares the cost-effectiveness of LQAS to a sub-district or district approach, by displaying the cost per intervention school that would be saved if LQAS was used, i.e. LQAS is more cost-effective where the plotted values are positive on the *y*-axis. For *S. mansoni*, a sub-district approach, using around three schools per sub-district, provided a slightly more cost-effective approach than LQAS in Western and Nyanza provinces, but was less cost-effective in Eastern province. Whilst a district approach appeared to achieve high levels of performance in Nyanza, Western and Eastern province (Figure 5.3), on several realizations, this approach failed to correctly

classify any intervention schools correctly which made it impossible to gain an estimate of cost-effectiveness. For a similar reason, sub-district/district cost-effectiveness results are not shown for Coast province. For STH, both sub-district and district approaches were more cost-effective than LQAS in all settings. The same was true for *P. falciparum* in Western and Nyanza provinces, however, cost-effectiveness for sub-district and district approaches in Eastern and Coast province could not be calculated as on several realisations these approaches completely failed to identify any intervention schools.

Across all species and settings, a sub-district approach was generally more cost-effective than a district approach. Furthermore, surveying around 3 schools per sub-district appeared to provide reasonable levels of cost-effectiveness across scenarios.
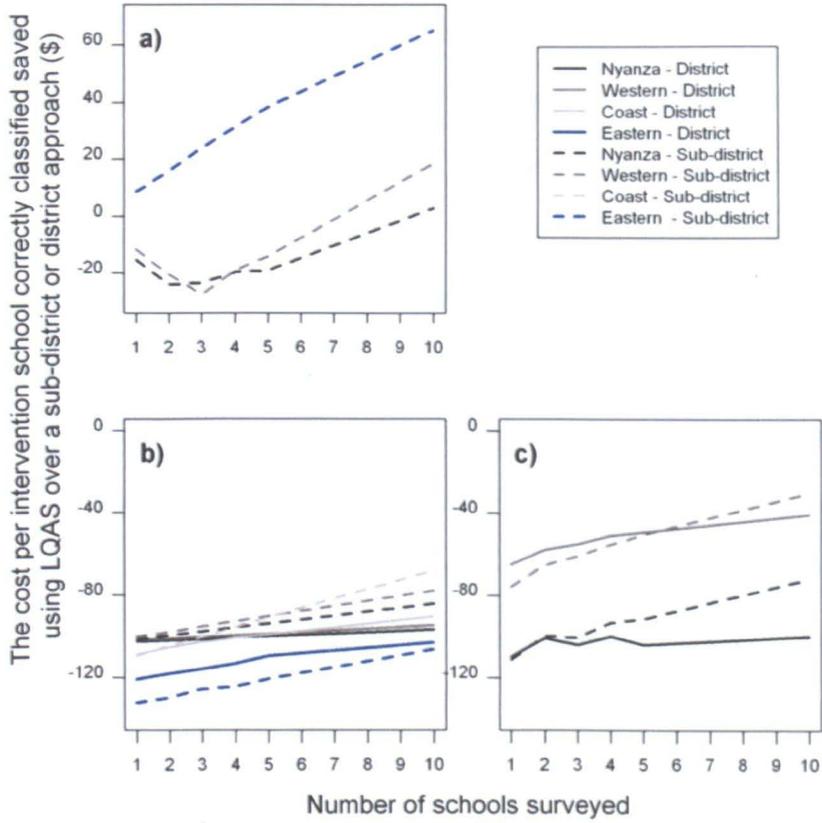
Figure 5.4 The difference in cost per intervention school correctly classified between sub-district/district approach and LQAS for a) *S. mansoni*, b) STH and c) *P. falciparum*. Note that results for a district approach for *S. mansoni* are not shown due to the fact that on several realizations, this approach completely failed to classify any schools. Similarly, results for a sub-district and district approach for *P. falciparum* in Coast and Eastern are not shown for the same reason.

Results of the cost sensitivity analysis are summarized in Figure 5.5 which compares the cost-effectiveness of LQAS to a sub-district/district approach, assuming three schools are surveyed per sub-district/district. It should be noted that as for the cost-effectiveness results shown in Figure 5.4, these results are only from moderate/high prevalence settings as sub-district/district approach did not achieve adequate performance in low prevalence settings. These results show that the cost-effectiveness of LQAS tends to increase if treatment costs are higher, or if three schools can be visited per day using LQAS. For *S. mansoni*, in these two alternative cost settings, LQAS becomes more cost effective than a sub-district approach. For STH, a sub-district and district approach remain more cost-effective than LQAS across all cost settings. For *P. falciparum*, a higher treatment cost results in LQAS becoming about as cost-effective as a sub-district approach, whereas LQAS remains less cost-effective than a sub-district/district approach even if three schools can be visited per day using LQAS.
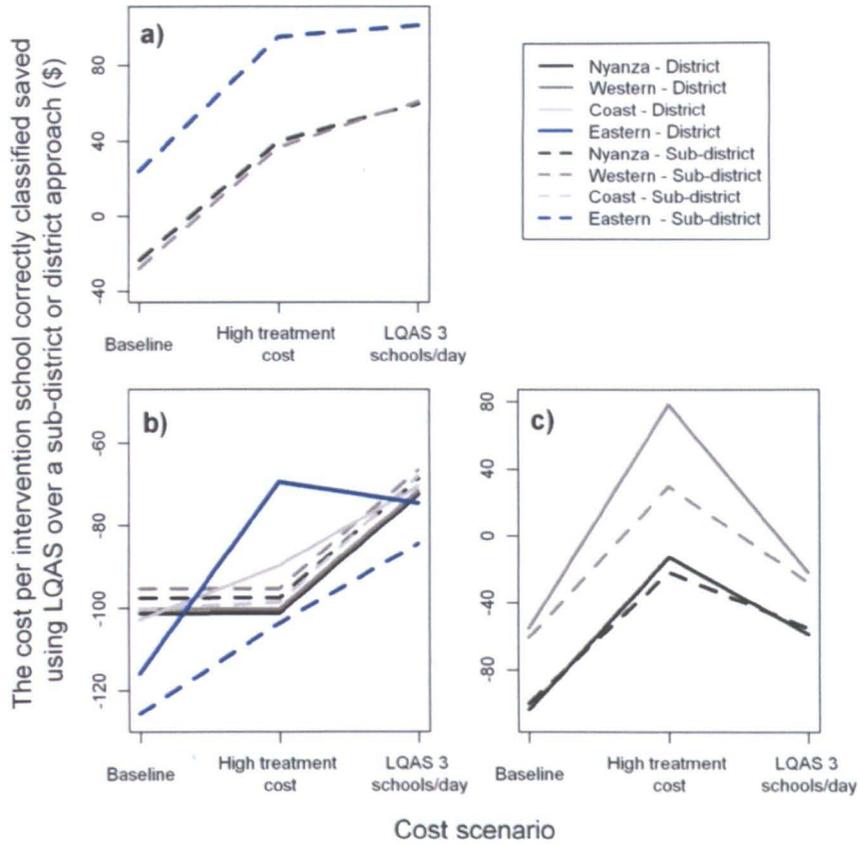
Figure 5.5 Results of the cost sensitivity analyses, comparing the cost-effectiveness of a sub-district/district approach (using 3 schools per sub-district/district) to that achieved using LQAS, for a) *S. mansoni*, b) STH and c) *P. falciparum*. Baseline assumes lower treatment costs of $0.295, $0.15 and $1.20 for *S. mansoni*, STH and *P. falciparum* respectively. High treatment cost assumes costs of $0.37, $0.39 and $5.10 for *S. mansoni*, STH and *P. falciparum*. Note that results for a district approach for *S. mansoni* are not shown due to the fact that on several realizations, this approach completely failed to classify any schools. Similarly, results for a sub-district and district approach for *P. falciparum* in Coast and Eastern are not shown for the same reason.

5.3.4 Implications for integrated survey designs

Whilst there is variation in performance and cost-effectiveness of each survey approach between different parasite, prevalence and province settings, three main conclusions are evident. Firstly, for *S. mansoni* and *P. falciparum*, in low prevalence settings (Coast - *S. mansoni*, Eastern and Coast - *P. falciparum*, Table 5.4), sub-district/district approaches are not appropriate as they fail to achieve acceptable levels of performance and cost-effectiveness (Figure 5.3). In such settings, LQAS should be used. Unfortunately it was not possible to investigate different survey approaches in settings of low STH prevalence (i.e. <20%), but it is likely that LQAS will become favourable as areas in which treatment is warranted become patchier. Further investigation on this topic is clearly required. Secondly, in areas of moderate/high prevalence, using 3 schools per sub-district resulted in relatively good levels of performance and cost-effectiveness for all species. Thirdly, assuming baseline survey and treatment costs, such a sub-district approach was more cost-effective than LQAS in almost every province for all three species (Figure 5.4). If, however, higher treatment costs were used, or it was assumed that three schools could be visited per day using LQAS, the benefit of a sub-district approach diminished across all species (Figure 5.5). For *S. mansoni* this resulted in LQAS becoming the more cost-effective approach. For STH, a sub-district/district approach remained the most cost-effective approach and for *P. falciparum* a sub-district approach became almost equally as cost-effective as LQAS (Figure 5.5).

From an integrated survey perspective, these results suggest that a two stage survey design might provide a balance between cost-effectiveness and operational ease. As a first stage, three schools per sub-district could be surveyed for all diseases. Within sub-districts of low *S. mansoni* or *P. falciparum* prevalence, a second LQAS stage could be carried out to help further guide

interventions. Although it is unclear exactly what the prevalence should be before a second stage

of LQAS is employed, a safe option would be to use the recommended treatment thresholds. For

example, following a first stage of sub-district level surveys, LQAS could be employed in sub-

districts where prevalence of S. mansoni is <10% and in sub-districts where P. falciparum

prevalence is <5%. Whilst the cost sensitivity analyses suggest that in some cases switching to

LQAS may be beneficial even in moderate prevalence settings, using the recommended treatment

thresholds would provide a balance between the considerable resource requirements of carrying

out LQAS over large scales and ensuring that intervention schools receive treatment. Figure 5.6

illustrates this approach to integrated mapping of these parasites.
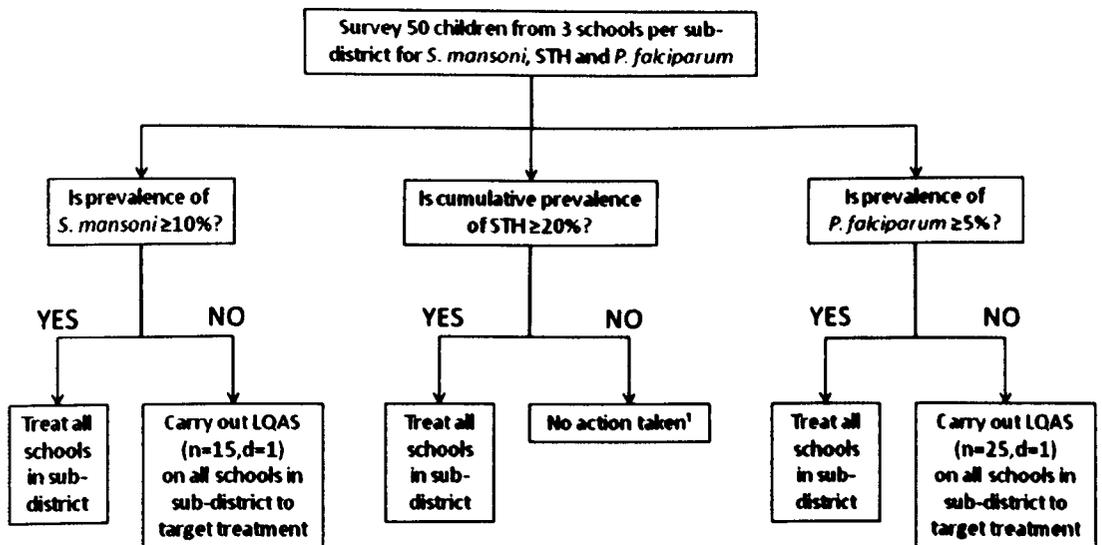
Figure 5.6 A decision tree for a possible integrated approach to mapping S. mansoni, STH and P. falciparum.
[1]Further work needed to show whether LQAS would be beneficial in settings where prevalence <20%.

5.4 DISCUSSION

There is increasing interest in integrated surveys for tropical diseases (Brooker and Utzinger, 2007; Brooker *et al.*, 2009a; Baker *et al.*, 2010). The importance of spatial scale in the design of such surveys and subsequent control activities is, however, not well understood. Here, computerized simulations were used to test the performance and cost-effectiveness of alternative survey approaches over different spatial scales for *S. mansoni*, STH and *P. falciparum*. Results suggest that for all parasites, a sub-district approach is generally more cost-effective than a district approach and that moving to finer resolution mapping using LQAS becomes more cost-effective in lower prevalence settings, in situations of high treatment cost or assuming three schools can be visited per day using LQAS. These results suggest that in this region, surveys for *S. mansoni*, STH and *P. falciparum* can be integrated at the sub-district level, using a random selection of three schools, and that in sub-districts with low prevalence of *S. mansoni* or *P. falciparum*, LQAS should be used to target interventions.

Results from the survey simulations have a number of implications for the design of independent surveys for *S. mansoni*, STH and *P. falciparum*. Firstly, for *S. mansoni*, LQAS is more cost-effective than a sub-district approach in low prevalence settings, in situations where treatment costs are high or when three schools can be visited per day using LQAS. This may be due, in part, to the focality of the disease which results in foci of infection which can be missed if a sub-district approach is adopted, particularly in low prevalence settings (Brooker, 2007). That said, in higher prevalence settings, a sub-district approach becomes an increasingly favourable option over LQAS as it becomes comparatively cost-effective and requires far less survey effort. A similar relationship was suggested by Brooker *et al.* (2005), who showed that in areas of very high

161

prevalence (>75%) of *S. mansoni* presumptive treatment without surveys may be a more cost-effective approach than targeted treatment using LQAS.

Secondly, for STH, a sub-district approach appears to be equally as cost-effective as a district approach over the settings investigated, although finer resolution mapping could be required in low prevalence settings, which were not investigated here. The fact that districts can be used for STH may, in part, be due to the fact that the dominant species, hookworm, appears to cluster over large scales (Brooker *et al.*, 2004b; Sturrock *et al.*, 2010). With this in mind, it should be possible to integrate surveys for STH with those for *S. mansoni* as carrying out surveys at the finer spatial resolution required for *S. mansoni* should provide adequate spatial resolution for STH. Furthermore, diagnosis of *S. mansoni* and STH both use Kato-Katz meaning that integrating STH surveys into those for *S. mansoni* only involves slightly more laboratory work. That said, antigen capture dipsticks, that detect schistosome CCA in urine, have shown excellent diagnostic performance for detection of *S. mansoni* infection in the field and may prove useful in the future (Stothard *et al.*, 2006; Legesse and Erko, 2007; Standley *et al.*, 2010).

Thirdly, for *P. falciparum*, sub-district and district approaches appear to be as cost-effective as each other over different prevalence settings. These approaches are, however, not suitable in low prevalence settings, as shown by their low performance in Coast and Eastern provinces where the simulated mean prevalence was 5.0% and 1.7% respectively. In such settings LQAS maintains high performance and cost-effectiveness. Furthermore, if the cost of intervention for *P. falciparum* is likely to be large (i.e. if bednets are distributed alongside IPT or several rounds of IPT are undertaken) then LQAS becomes the most cost-effective approach, even in moderate/high

162

prevalence settings. In practice, LQAS appears to be relatively sparsely used for *P. falciparum* surveys. In Mozambique, it has been used to provide regional information on bednet distribution using data collected during national MIS (Biedron *et al.*, 2010). In terms of disease prevalence estimates, LQAS has also been used to identify schools with a prevalence of ≥15% using a sampling plan of (36,2) (Rabarijaona *et al.*, 2001). Results from this study show that LQAS may also prove useful for detecting schools where prevalence is ≥5%.

Results of the cost sensitivity analysis show that generally, the benefit, in terms of cost-effectiveness, of using a sub-district approach over LQAS appears to diminish if treatment costs are higher or if three schools can be visited per day using LQAS. This is due to the fact that if treatment costs are higher, more accurate survey methods, such as LQAS, become more cost-effective as they minimize the cost associated with administration of unwarranted treatment. Equally, if three schools per day can be visited using LQAS, and therefore total survey costs are lower, then it is natural that the cost-effectiveness of LQAS will improve. Despite the increase in comparative cost-effectiveness of LQAS in these settings, only in the case of *S. mansoni* did this make LQAS the most cost-effective approach across settings. For both STH and *P. falciparum*, a sub-district/district approach generally remained more cost-effective than LQAS.

From an integrated survey perspective, these results suggest that a two stage integrated survey design would help to maximize cost-effectiveness and operational ease. As a first stage, surveying three schools per sub-district provides good levels of performance and cost-effectiveness for all species. In sub-districts that do not qualify for treatment, a second stage of finer resolution mapping using LQAS could be used to target interventions. Further theoretical and practical work

to explore this approach is warranted. The obvious effect of prevalence on the choice of survey approach employed also has implications for long term control. As programmes begin to expand control activities and prevalence decreases, finer resolution mapping may be required to ensure that persistent foci of infection are not missed.

There are a number of limitations of the current study that merit mention. Firstly, it is evident that a multitude of factors affect the performance and cost-effectiveness of survey designs. Similar studies in settings with varying spatial characteristics, survey and treatment costs, prevalence and district and sub-district sizes are therefore required before extrapolations of conclusions are made elsewhere. Such analyses would also help to further untangle the various influences these factors have on the performance of the different survey designs.

Secondly, conclusions have been made about integrated survey designs, whilst survey costs were considered independently for each parasite. In reality, survey costs for any integrated survey design would be shared between the three species. An evaluation of the proposed two stage integrated survey design would be useful to formally test its performance and cost-effectiveness.

Thirdly, the fact that an underlying large scale trend could not be removed from the *P. falciparum* data may affect the suitability of the conditionally simulated data to act as a gold standard. The reason for this is unclear, but may be due to slight differences in prevalence over time which when combined with a survey conducted in different places at different times, may result in a spatial trend which doesn't follow an obvious relationship with longitude, latitude or land surface

temperature. Indeed, the data used for these analyses were collected in three separate periods between September 2008 and March 2010. The existence of this trend may mean that in reality, infection is more focal than described here. If this is the case, however, evidence from the results of simulations with *S. mansoni* infection suggests that conducting surveys at a sub-district level using 3 schools still provides a robust method of classification in more focal settings. Repeating the analyses on predictive risk surfaces generated using Bayesian models which incorporate more covariate information, including time of survey, may help to resolve this issue and would further our understanding of the importance of spatial heterogeneity on the performance of these survey approaches.

Fourthly, it should be remembered that this study is assuming the use of school children is an appropriate target survey group, which is based on epidemiological and operational considerations. In settings of low *P. falciparum* prevalence, infection is likely to be more evenly distributed over different age groups, due to patterns of acquired immunity (Smith et al., 2007a). In these situations, LQAS at schools may not be sufficient to identify those individuals contributing most to transmission, and further surveys in communities may be required. Tracing infected children back to their communities may provide a solution and warrants further investigation.

Fifthly, the assumption was made that control programmes are balancing the cost of surveys against the cost of misclassification in the form of administration of unwarranted treatment. In reality, minimizing the number of drugs administered unnecessarily may not be a priority, particularly if the drugs are donated or maximizing the number of drugs distributed is a seen to be a priority by funding bodies.

Finally, only the use of LQAS to map these diseases at a fine scale was made. As chapter 4 has described, there may be an opportunity to use geostatistical methods to target treatment at a fine scale using only a subset of schools. Further work investigating geostatistical methods to optimize surveys for STH and *P. falciparum* would be interesting. Additionally, a number of online atlases now exist for these diseases such as GAHI (Brooker *et al.*, 2009b) and the Malaria Atlas Project (Hay *et al.*, 2009). Exploring survey designs that are able to 'fill in the gaps' of such atlases may help to reduce the survey effort required to produce fine scale information. Despite these limitations, these results strongly suggest that a sub-district approach is an appropriate scale at which to conduct integrated surveys for *S. mansoni*, STH and *P. falciparum* and that finer resolution mapping becomes more cost-effective in lower prevalence settings.

In summary, this study has used unique computerized simulations to evaluate the performance and cost-effectiveness of carrying out surveys and treatment at varying spatial scales for *S. mansoni*, STH and *P. falciparum* in Kenya. Results suggest that for all diseases, surveying three schools per sub-district provides good levels of performance and cost-effectiveness, but in areas of lower prevalence, finer resolution mapping is required. From an integrated survey standpoint, these results suggest that a two stage design, whereby LQAS is employed in low prevalence sub-districts, would provide a balance between cost-effectiveness and operational ease. These results also suggest that finer resolution mapping of infection will be required as prevalence drops as a result of control activities. Further work using data from a range of transmission and ecological settings would be beneficial.

CHAPTER 6 | Summary and Discussion of Findings

6.1 Summary of Findings

The overall aim of this thesis was to rigorously examine survey methods for STH, *S. mansoni*, and *P. falciparum*, accounting for spatial heterogeneity and cost, with the intention of devising an integrated approach to conducting surveys. The justification of concentrating on these diseases arises from their considerable spatial overlap as well as the fact that for all infections school children are an epidemiologically and operationally suitable sample population (Brooker *et al.*, 2006c; Brooker *et al.*, 2009c; Gitonga *et al.*, 2010). The introduction gave an overview of the biological and epidemiological features of these parasitic infections which influence the way in which they are surveyed for, current survey and rapid assessment methods used and opportunities and challenges for integrated survey designs. Key areas of research that were identified included a better appreciation of both spatial aspects of diseases and cost and their implications for optimal survey designs. More specifically, it was recognised that there are currently no rapid survey methods for STH, that current rapid survey methods for *S. mansoni* are operationally challenging over large scales and that there is a need to explore survey methods for malaria using updated intervention thresholds. The following chapters aimed to address these issues and to provide a framework for integrated surveys.

As a first step, chapter 2 described two recent attempts to design and carry out integrated surveys in East Africa. These studies attempted to integrate separate survey designs recommended by WHO into a single integrated approach. Integrated surveys were more complex in Southern Sudan than in Kenya due to the fact that different combinations of diseases were surveyed for in different villages and, due to the inclusion of LF, individuals of different ages were surveyed within those villages. In addition to highlighting the practical and operational challenges of planning and implementing integrated surveys, these experiences emphasised the need for a better understanding of the influence of spatial heterogeneity and cost on the choice of survey design.

For STH, current survey recommendations are based on the use of ecological zones, which are often not easy to define and rarely used in practice. Furthermore, results based on such an approach are often difficult to translate to the appropriate administrative level (typically the district) used to implement control interventions. Chapter 3 aimed to address these issues, using computerized sampling simulations in East Africa to investigate the performance and cost-effectiveness of different sized school based survey designs that aim to classify districts according to WHO treatment thresholds. Results showed that, over a number of settings, surveying small numbers of children in four to five randomly selected schools per district provides a rapid and cost-effective approach for the delivery of albendazole at the district level.

For *S. mansoni*, WHO currently recommends the use of schools or sub-districts as IUs due to the focal nature of infection. Previous work has shown that LQAS can be used as a rapid and cost-effective method to classify schools according to treatment thresholds (Brooker et al., 2005). Such a method, however, becomes impractical over large scales and alternative survey designs are

required. Chapter 4 explored the use of geostatistics to help reduce the number of survey sites required, using a grid of points, with some close pairs, to enable spatial prediction of prevalence over a study region. Simulations showed that whilst LQAS is able to correctly classify a higher proportion of schools that qualify for treatment, a geostatistical approach, using a grid of around 13.5km with 50 close pairs of points in combination with spatial interpolation, provides a more cost-effective method to identify schools that qualify for treatment. Such a geostatistical method is computationally complex, but in situations where surveys are to be conducted over large scales, and where technical resources allow, it may offer an alternative to LQAS.

Both chapters 3 and 4 make the assumption that the current IUs of district and school, for STH and *S. mansoni* respectively, are optimal scales at which to conduct control programmes. The empirical basis for these scales is, however, not well defined. Chapter 5 aimed to explore the implications of mapping and carrying out control activities at varying spatial scales for STH, *S. mansoni* and *P. falciparum*, and sought to assess whether there is a scale at which surveys for these parasites can be validly integrated into a single design. Results showed that a sub-district approach, using three randomly selected schools per sub-district, provides good levels of performance and cost-effectiveness across species and settings. Furthermore, it was shown that in settings of low *S. mansoni* or *P. falciparum* prevalence, higher resolution mapping is required. Taken together, these results suggest that in East Africa, a two-stage survey design, whereby LQAS is carried out in sub-districts that do not qualify for treatment, may provide a robust approach to integrated mapping of these diseases. As shown in chapter 4 for *S. mansoni*, where technical resources allow, such high resolution mapping may be achieved using geostatistical methods, as opposed to LQAS.

6.2 FUTURE DIRECTIONS

Survey designs for tropical diseases have received relatively little attention in the literature, despite the fact that they form the foundations of control efforts. As we move into a period of integrated control of tropical diseases, further research into methods that allow distributions and overlaps of disease to be understood are required. In the last six months, WHO-AFRO have developed a draft protocol for integrated mapping of LF, schistosomiasis, STH, onchocerciasis and trachoma (Likezo Mubila, pers. comm), however, these have yet to be validated in the field. These guidelines are, however, essentially a combination of current separate survey methods into a single framework. I would argue that there is an opportunity to further refine survey designs by considering a number of factors including spatial heterogeneity, prevalence and survey and treatment costs, which, as this thesis has shown, all influence the spatial resolution of surveys and the sampling effort required.

There are a number of opportunities for extending and improving the research covered in this thesis. The use of conditionally simulated data provides realistic fully enumerated gold standard data, which would otherwise be impossible to obtain. Paradoxically, this approach is the thesis' main drawback, as the context specific nature of the gold standard data makes extrapolation of conclusions to other settings difficult. The advent of Bayesian geostatistical predictive maps, however, provides an opportunity to carry out sampling simulations using 'gold standard' data across various prevalence and geographic settings (Hay *et al.*, 2009; Magalhães *et al.*, 2011). Using gold standard data from non-stationary Bayesian geostatistical predictive maps will allow a better understanding of the influence of spatial heterogeneity on survey design (Beck-Worner *et al.*, 2007; Gosoniu *et al.*, 2009).

One potentially exciting extension of this work would be to incorporate survey simulations into mathematical intervention models (Chan and Bundy, 1997; Griffin *et al.*, 2010). Incorporating an initial survey step in intervention models would allow an exploration of the implications of the performance of different survey designs - and hence the proportion of infected individuals that would be correctly treated based on that design - on disease transmission. Or thinking backwards, allow us to elucidate the most cost-effective long term intervention strategy and the level of performance required by surveys to achieve that strategy. It would also allow an examination of whether changes in surveillance strategies are required as prevalence decreases over time due to control efforts.

It would also be interesting to incorporate diagnostic performance into these sampling simulations, as this thesis has essentially assumed 100% sensitivity and specificity during sampling which is unrealistic (Booth *et al.*, 2003; Murray *et al.*, 2008). That said, the data on which simulations are based, were collected using Kato-Katz or malaria RDT diagnosis. In addition to including diagnostic performance during the sampling part of the simulations, therefore, the gold standard data would also have to be adjusted to reflect the inherent error in the original survey data. A consideration of diagnostic performance is likely to become more important in lower prevalence settings, where infected individuals can be easily missed using conventional methods (Okell *et al.*, 2009).

Another possible avenue of research is to build in covariate data into survey design. A tremendous amount of information relating to disease risk is available to the scientific community, such as remotely sensed data (Hay *et al.*, 2006b). Using this wealth of covariate data in geostatistical

171

predictive models can help estimate prevalence. Optimizing survey designs for (geo)statistical model building could considerably reduce the amount of survey effort required. The use of covariate information in survey design has been an area of interest in ecology (Hirzel and Guisan, 2002) but has yet to be fully explored in a disease setting. On a related note, Bayesian geostatistical risk maps have been produced for a number of diseases, providing prevalence estimates with a varying degree of certainty across SSA. Devising survey approaches that 'fill in the gaps' in areas where model uncertainty is high will be another crucial area of research. Survey designs that minimize space/time kriging variance may be a possible approach and warrant further investigation.

This thesis has shown that the design of optimal survey methods requires a consideration of the spatial heterogeneity and prevalence of parasite species as well as the costs of the surveys and intervention programme. Based on the analyses, a two stage framework for integrated surveys that allows for flexibility according to parasite species and their transmission settings is proposed. Validating this framework over a range of settings, and incorporating diagnostic performance and transmission models, would add substantial value to this work. This is particularly pertinent as we move into an era of renewed interest in tropical diseases where reliable and cost-effective survey methods are required if the considerable challenge of disease control and elimination is to be met.

The attached CD contains examples of bespoke script used in the thesis as well as some example datasets. Each folder contains the relevant data used in the script file for that folder, hence some datasets are in more than one folder. The following files are included:

Conditional Simulation folder

ConditionalSimulationSTH.R – This is the R script used to conditionally simulate realizations of STH species. Coast province is used as an example.

gok_primary_schools_coast.txt – A text file with the coordinates and district code for all public primary schools in Coast province.

kenya_STH_rawprev.txt – A text file showing the coordinates and raw prevalence values for STH infection in Kenya.

S. mansoni folder

EndClassEst.09_50.R – This is the R script used to implement the Lattice plus Close Pairs design for *S. mansoni* sampling in Western and Nyanza provinces in Kenya, using a grid size of .09 decimal degrees as an example.

PosteriorSamplesSmansWesternNyanza.csv - Dataset of conditionally simulated realizations of *S. mansoni* for Western and Nyanza provinces, Kenya. Each column is a single realisation.

Prim_schools_western_nyanza_mask.csv – A list of those schools included in the ecological mask.

Smans_primary_schools_western_nyanza.csv – Coordinates of all public primary schools in Western and Nyanza provinces, Kenya.

Scale and Sampling folder

Smans_scale_sampling_coast.R – This is the script that compares sub-district to district sampling, using *S. mansoni* in Coast province as an example.

SmansPosteriorSamplesCoast.csv –Dataset of conditionally simulated realizations of *S. mansoni* for Coast province, Kenya. Each column is a single realisation.

SmansPrimarySchoolsCoast.csv - Dataset of all public primary schools in Coast province with coordinates (X and Y), district code (DIST), sub district code (SDCODE), district area (AREA_D) and sub district area (AREA_SD).

SurvCostDistrictCoast.csv – The estimated cost in US$ of conducting surveys at the district level in Coast province using 1, 2, 3, 4, 5, or 10 schools per district.

SurvCostSubDistrictCoast.csv – The estimated cost in US$ of conducting surveys at the sub-district level in Coast province using 1, 2, 3, 4, 5, or 10 schools per sub-district.

STH folder

DistrictSamplingSTH.R – This is the script used to simulate district sampling of STH in Coast province.

gok_primary_schools_coast.txt – A text file with the coordinates and district code for all public primary schools in Coast province.

PosteriorSamplesCoast.csv - Dataset of conditionally simulated realizations of hookworm for Coast province, Kenya. Each column is a single realisation.

PosteriorSamplesAscCoast.csv - Dataset of conditionally simulated realizations of *A. lumbricoides* for Coast province, Kenya. Each column is a single realisation.

PosteriorSamplesTriCoast.csv - Dataset of conditionally simulated realizations of *T. trichiura* for Coast province, Kenya. Each column is a single realisation.

LQAS folder

LQAS_Smans.R – This is the R script that runs simulations of Lot Quality Assurance Sampling using *S. mansoni* in Coast province, Kenya as an example

SmansPosteriorSamplesCoast.csv –Dataset of conditionally simulated realizations of *S. mansoni* for Coast province, Kenya. Each column is a single realisation.

SmansPrimarySchoolsCoast.csv - Dataset of all public primary schools in Coast province with coordinates (X and Y), district code (DIST), sub district code (SDCODE), district area (AREA_D) and sub istrict area (AREA_SD).

REFERENCES

Abramo, C., Fontes, C.J., Krettli, A.U., (1995). Cross-reactivity between antibodies in the sera of individuals with leishmaniasis, toxoplasmosis, and Chagas' disease and antigens of the blood-stage forms of Plasmodium falciparum determined by indirect immunofluorescence. *American Journal of Tropical Medicine and Hygiene* 53 (2), 202-205.

Alexander, N.D., Moyeed, R.A., Hyun, P.J., Dimber, Z.B., Bockarie, M.J., Stander, J., Grenfell, B.T., Kazura, J.W., Alpers, M.P., (2003). Spatial variation of *Anopheles*-transmitted *Wuchereria bancrofti* and *Plasmodium falciparum* infection densities in Papua New Guinea. *Filaria Journal* 2 (1), 14.

Anderson, R.M., May, R.M., (1979). Prevalence of Schistosome infections within molluscan populations: observed patterns and theoretical predictions. *Parasitology* 79 (1), 63-94.

Anderson, R.M., May, R.M., (1985). Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology* 24, 1-101.

Anderson, R.M., (1986). The population dynamics and epidemiology of intestinal nematode infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80 (5), 686-596.

Anderson, R.M., May, R.M., (1991). Infectious Diseases of Humans: dynamics and control. Oxford University Press, Oxford.

Anker, M., (1991). Epidemiological and statistical methods for rapid health assessment: introduction. *World Health Statistics Quarterly* 44 (0379-8070 (Print)), 94-97.

Ansell, J., Guyatt, H.L., (2002). Comparative cost-effectiveness of diagnostic tests for urinary schistosomiasis and the implications for school health programmes. *Annals of Tropical Medicine and Parasitology* 96 (2), 145-153.

Arap Siongok, T.K., Mahmoud, A.A., Ouma, J.H., Warren, K.S., Muller, A.S., Handa, A.K., Houser, H.B., (1976). Morbidity in Schistosomiasis mansoni in relation to intensity of infection: study of a community in Machakos, Kenya. *American Journal of Tropical Medicine and Hygiene* 25 (2), 273-284.

Ashton, R., Kefyalew, T., Tesfaye, G., Pullan, R., Yadeta, D., Reithinger, R., Kolaczinski, J., Brooker, S., (2011). School-based surveys of malaria in Oromia Regional State, Ethiopia: a rapid survey method for malaria in low transmission settings. *Malaria Journal* 10 (1), 25.

Ayele, B., Erko, B., Legesse, M., Hailu, A., Medhin, G., (2008). Evaluation of circulating cathodic antigen (CCA) strip for diagnosis of urinary schistosomiasis in Hassoba school children, Afar, Ethiopia. *Parasite* 15 (1), 69-75.

Baker, M.C., Mathieu, E., Fleming, F.M., Deming, M., King, J.D., Garba, A., Koroma, J.B., Bockarie, M., Kabore, A., Sankara, D.P., Molyneux, D.H., (2010). Mapping, monitoring, and surveillance of neglected tropical diseases: towards a policy framework. *The Lancet* 375 (9710), 231-238.

Beaver, P.C., (1953). Persistence of hookworm larvae in soil. *American Journal of Tropical Medicine and Hygiene* 2 (1), 102-108.

Beck-Worner, C., Raso, G., Vounatsou, P., N'Goran, E.K., Rigo, G., Parlow, E., Utzinger, J., (2007). Bayesian spatial risk prediction of *Schistosoma mansoni* infection in western Cote d'Ivoire using a remotely-sensed digital elevation model. *American Journal of Tropical Medicine and Hygiene* 76 (5), 956-963.

Bejon, P., Williams, T.N., Liljander, A., Noor, A.M., Wambua, J., Ogada, E., Olotu, A., Osier, F.H.A., Hay, S.I., FÃ¤rnert, A., Marsh, K., (2010). Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. *PLoS Medicine* 7 (7), e1000304.

Bennett, S., Radalowicz, A., Vella, V., Tomkins, A., (1994). A computer simulation of household sampling schemes for health surveys in developing countries. *International Journal of Epidemiology* 23 (6), 1282-1291.

Bergquist, R., Johansen, M.V., Utzinger, J., (2009). Diagnostic dilemmas in helminthology: what tools to use and when? *Trends in Parasitology* 25 (4), 151-156.

Biedron, C., Pagano, M., Hedt, B.L., Kilian, A., Ratcliffe, A., Mabunda, S., Valadez, J.J., (2010). An assessment of Lot Quality Assurance Sampling to evaluate malaria outcome indicators: extending malaria indicator surveys. *International Journal of Epidemiology* 39 (1), 72-79.

Booth, M., Bundy, D.A., (1995). Estimating the number of multiple-species geohelminth infections in human communities. *Parasitology* 111 (Pt 5), 645-653.

Booth, M., Bundy, D.A., Albonico, M., Chwaya, H.M., Alawi, K.S., Savioli, L., (1998). Associations among multiple geohelminth species infections in schoolchildren from Pemba Island. *Parasitology* 116 (Pt 1), 85-93.

Booth, M., Vounatsou, P., N'Goran E, K., Tanner, M., Utzinger, J., (2003). The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Cote d'Ivoire. *Parasitology* 127 (Pt 6), 525-531.

Brady, M.A., Hooper, P.J., Ottesen, E.A., (2006). Projected benefits from integrating NTD programs in sub-Saharan Africa. *Trends in Parasitology* 22 (7), 285-291.

Brooker, S., Michael, E., (2000). The potential of geographical information systems and remote sensing in the epidemiology and control of human helminth infections. *Advances in Parasitology* 47, 245-288.

Brooker, S., Rowlands, M., Haller, L., Savioli, L., Bundy, D.A.P., (2000). Towards an atlas of human helminth infection in sub-Saharan Africa: the use of Geographical Information Systems (GIS). *Parasitology Today* 16 (7), 303-307.

Brooker, S., Miguel, E.A., Waswa, P., Namunyu, R., Moulin, S., Guyatt, H., Bundy, D.A.P., (2001). The potential of rapid screening methods for *Schistosoma mansoni* in western Kenya. *Annals of Tropical Medicine and Parasitology* 95 (4), 343-351.

Brooker, S., Hay, S.I., Tchuem Tchuente, L.A., Ratard, R., (2002). Using NOAA-AVHRR data to model human helminth distributions in planning disease control in Cameroon, West Africa. *Photogrammetric Engineering & Remote Sensing* 68, 175-179.

Brooker, S., Clarke, S., Njagi, J.K., Polack, S., Mugo, B., Estambale, B., Muchiri, E., Magnussen, P., Cox, J., (2004a). Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Medicine & International Health* 9 (7), 757-766.

Brooker, S., Kabatereine, N.B., Tukahebwa, E.M., Kazibwe, F., (2004b). Spatial analysis of the distribution of intestinal nematode infections in Uganda. *Epidemiology and Infection* 132 (6), 1065-1071.

Brooker, S., Kabatereine, N.B., Myatt, M., Stothard, R.J., Fenwick, A., (2005). Rapid assessment of *Schistosoma mansoni*: the validity, applicability and cost-effectiveness of the Lot Quality Assurance Sampling method in Uganda. *Tropical Medicine and International Health* 10 (7), 647-658.

Brooker, S., Alexander, N., Geiger, S., Moyeed, R.A., Stander, J., Fleming, F., Hotez, P.J., Correa-Oliveira, R., Bethony, J., (2006a). Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *International Journal for Parasitology* 36 (10-11), 1143-1151.

Brooker, S., Clements, A.C., Bundy, D.A.P., (2006b). Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology* 62, 221 - 261.

Brooker, S., Clements, A.C., Hotez, P.J., Hay, S.I., Tatem, A.J., Bundy, D.A., Snow, R.W., (2006c). The co-distribution of *Plasmodium falciparum* and hookworm among African schoolchildren. *Malaria Journal* 5 (1), 99.

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., Akhwale, W.S., Pullan, R., Estambale, B., Clarke, S., Hotez, P.J., (2007). Epidemiology of *Plasmodium*-Helminth coinfeciton in Africa: potential impact on anaemia and prospects for combining control *American Journal of Tropical Medicine and Hygiene* 77 (6 Suppl), 88-98.

Brooker, S., Utzinger, J., (2007). Integrated disease mapping in a polyparasitic world. *Geospatial health* 1 (2), 141-146.

Brooker, S., Bundy, D.A.P., (2008). Soil-transmitted helminths (geohelminths). In: *Manson's Tropical Diseases* (22nd Edition), edited by Cook, G.C.,Zumla, A.I., Elsevier, London. pp. 1515-1548.

Brooker, S., Hotez, P.J., Bundy, D.A.P., (2008a). Hookworm-related anaemia among pregnant women: a systematic review. *PLoS Neglected Tropical Diseases* 2 (9), e291.

Brooker, S., Kabatereine, N.B., Fleming, F., Devlin, N., (2008b). Cost and cost-effectiveness of nationwide school-based helminth control in Uganda: intra-country variation and effects of scaling-up. *Health Policy and Planning* 23 (1), 24-35.

Brooker, S., Kabatereine, N.B., Gyapong, J.O., Stothard, J.R., Utzinger, J., (2009a). Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology* 136 (13), 1707-1718.

Brooker, S., Kabatereine, N.B., Smith, J.L., Mupfasoni, D., Mwanje, M.T., Ndayishimiye, O., Lwambo, N.J., Mbotha, D., Karanja, P., Mwandawiro, C., Muchiri, E., Clements, A.C., Bundy, D.A., Snow, R.W., (2009b). An updated atlas of human helminth infections: the example of East Africa. *International Journal of Health Geographics* 8, 42.

Brooker, S., Kolaczinski, J.H., Gitonga, C.W., Noor, A.M., Snow, R.W., (2009c). The use of schools for malaria surveillance and programme evaluation in Africa. *Malaria Journal* 8 (1), 231.

Brooker, S., (2010). Estimating the global distribution and disease burden of intestinal nematode infections: Adding up the numbers – A review. *International Journal for Parasitology* 40 (10), 1137–1144.

Brooker, S., Hotez, P.J., Bundy, D.A.P., (2010). The global atlas of helminth infection: mapping the way forward in neglected tropical disease control. *PLoS Neglected Tropical Diseases* 4 (7), 1137-1144.

Bundy, D.A.P., (1990). Control of intestinal nematode infections by chemotherapy: mass treatment versus diagnostic screening. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 84 (5), 622-625.

Burnham, G., (1998). Onchocerciasis. *The Lancet* 351 (9112), 1341-1346.

Burton, M.J., Mabey, D.C.W., (2009). The global burden of trachoma: a review. *PLoS Neglected Tropical Diseases* 3 (10), e460.

Carabin, H., Guyatt, H., Engels, D., (2000). A comparative analysis of the cost-effectiveness of treatment based on parasitological and symptomatic screening for *Schistosoma mansoni* in Burundi. *Tropical Medicine and International Health* 5 (3), 192-202.

Carter, R., Mendis, K.N., Roberts, D., (2000). Spatial targeting of interventions against malaria. *Bulletin of the World Health Organisation* 78 (12), 1401-1411.

Cecchi, G., Paone, M., Franco, J.R., Fevre, E.M., Diarra, A., Ruiz, J.A., Mattioli, R.C., Simarro, P.P., (2009). Towards the Atlas of human African trypanosomiasis. *International Journal of Health Geographics* 8, 15.

Chan, M.S., Bundy, D.A.P., (1997). Modelling the dynamic effects of community chemotherapy on patterns of morbidity due to *Schistosoma mansoni*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 91 (2), 216-220.

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

Clarke, S.E., Jukes, M.C., Njagi, J.K., Khasakhala, L., Cundill, B., Otido, J., Crudder, C., Estambale, B.B., Brooker, S., (2008). Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomised, double-blind, placebo-controlled trial. *The Lancet* 372 (9633), 127-138.

Clements, A.C., Lwambo, N.J., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J.P., Fenwick, A., Brooker, S., (2006a). Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine and International Health* 11 (4), 490-503.

Clements, A.C., Moyeed, R., Brooker, S., (2006b). Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* 133 (6), 711-719.

Clements, A.C., Brooker, S., Nyandindi, U., Fenwick, A., Blair, L., (2008a). Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *International Journal for Parasitology* 38 (3-4), 401-415.

Clements, A.C., Garba, A., Sacko, M., Toure, S., Dembele, R., Landoure, A., Bosque-Oliva, E., Gabrielli, A.F., Fenwick, A., (2008b). Mapping the probability of schistosomiasis and associated uncertainty, West Africa. *Emerging Infectious Diseases* 14 (10), 1629-1632.

Clements, A.C., Bosque-Oliva, E., Sacko, M., Landoure, A., Dembele, R., Traore, M., Coulibaly, G., Gabrielli, A.F., 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., Deville, M.A., Ndayishimiye, O., Brooker, S., Fenwick, A., (2010a). Spatial co-distribution of neglected tropical diseases in the east African great lakes region: revisiting the justification for integrated control. *Tropical Medicine and International Health* 15 (2), 198-207.

Clements, A.C., Kur, L.W., Gatpan, G., Ngondi, J.M., Emerson, P.M., Lado, M., Sabasio, A., Kolaczinski, J.H., (2010b). Targeting trachoma control through risk mapping: the example of Southern Sudan. *PLoS Neglected Tropical Diseases* 4 (8), e799.

Clennon, J.A., King, C.H., Muchiri, E.M., Karuiki, H.C., Ouma, J.H., Mungai, P., Kitron, U., (2004). Spatial patterns of urinary schistosomiasis infection in a highly endemic area of coastal Kenya. *American Journal of Tropical Medicine and Hygiene* 70 (4), 443-448.

Connelly, L.B., (2003). Balancing the number and size of sites: an economic approach to the optimal design of cluster samples. *Controlled Clinical Trials* 24 (5), 544-559.

Corran, P., Coleman, P., Riley, E., Drakeley, C., (2007). Serology: a robust indicator of malaria transmission intensity? *Trends in Parasitology* 23 (12), 575-582.

Craig, M.H., Snow, R.W., le Sueur, D., (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today* 15 (3), 105-111.

Crainiceanu, C.M., Diggle, P.J., Rowlingson, B., (2008). Bivariate binomial spatial modeling of *Loa loa* prevalence in tropical Africa. *Journal of the American Statistical Association* 103, 21-37.

Cressie, N., (2000). Geostatistical methods for mapping environmental exposures. In: *Spatial Epidemiology, Methods and Applications* edited by Elliott, P., Wakefield, J.C., Best, N.G.,Briggs, D.J., Oxford University Press, Oxford. pp. 185-204.

Crompton, D.W., Nesheim, M.C., (2002). Nutritional impact of intestinal helminthiasis during the human life cycle. *Annual Review of Nutrition* 22, 35-59.

de Silva, N.R., Brooker, S., Hotez, P.J., Montresor, A., Engels, D., Savioli, L., (2003). Soil-transmitted helminth infections: updating the global picture. *Trends in Parasitology* 19 (12), 547-551.

De Sole, G., Baker, R., Dadzie, K.Y., Giese, J., Guillet, P., Keita, F.M., (1991a). Onchocerciasis distribution and severity in five West African countries. *Bulletin of the World Health Organisation* 69 (6), 689-697.

De Sole, G., Giese, J., Keita, F.M., Remme, J., (1991b). Detailed epidemiological mapping of three onchocerciasis foci in West Africa. *Acta Tropica* 48 (3), 203-213.

Deganello, R., Cruciani, M., Beltramello, C., Duncan, O., Oyugi, V., Montresor, A., (2007). *Schistosoma hematobium* and *S. mansoni* among children, Southern Sudan. *Emerging Infectious Diseases* 13 (10), 1504-1506.

Deitchler, M., Valadez, J.J., Egge, K., Fernandez, S., Hennigan, M., (2007). A field test of three LQAS designs to assess the prevalence of acute malnutrition. *International Journal of Epidemiology* 36 (4), 858-864.

Deitchler, M., Deconinck, H., Bergeron, G., (2008). Precision, time, and cost: a comparison of three sampling designs in an emergency setting. *Emerging Themes in Epidemiology* 5 (6).

Diggle, P.J., Lophaven, S., (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics* 33 (1), 53-64.

Diggle, P.J., Ribeiro Jnr, P.J., (2007). Model-based geostatistics. Springer, New York.

Diggle, P.J., Thomson, M.C., Christensen, O.F., Rowlingson, B., Obsomer, V., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Kamgno, J., Remme, J.H., Boussinesq, M., Molyneux, D.H., (2007). Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty. *Annals of Tropical Medicine and Parasitology* 101 (6), 499-509.

Dodd, R., Cassels, A., (2006). Health, development and the Millennium Development Goals. *Annals of Tropical Medicine and Parasitology* 100 (5-6), 379-387.

Drakeley, C., Schellenberg, D., Kihonda, J., Sousa, C.A., Arez, A.P., Lopes, D., Lines, J., Mshinda, H., Lengeler, C., Armstrong Schellenberg, J., Tanner, M., Alonso, P., (2003). An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. *Tropical Medicine and International Health* 8 (9), 767-774.

Drakeley, C.J., Corran, P.H., Coleman, P.G., Tongren, J.E., McDonald, S.L., Carneiro, I., Malima, R., Lusingu, J., Manjurano, A., Nkya, W.M., Lemnge, M.M., Cox, J., Reyburn, H., Riley, E.M., (2005). Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* 102 (14), 5108-5113.

Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, B.J., Stoddart, G.L., (2005). Methods for the economic evaluation of health care programmes. Oxford University Press, Oxford.

Eigege, A., Richards, F.O., Jr., Blaney, D.D., Miri, E.S., Gontor, I., Ogah, G., Umaru, J., Jinadu, M.Y., Mathai, W., Amadiegwu, S., Hopkins, D.R., (2003). Rapid assessment for lymphatic filariasis in central Nigeria: a comparison of the immunochromatographic card test and hydrocele rates in an area of high endemicity. *American Journal of Tropical Medicine and Hygiene* 68 (6), 643-646.

Emerson, P.M., Ngondi, J., Biru, E., Graves, P.M., Ejigsemahu, Y., Gebre, T., Endeshaw, T., Genet, A., Mosher, A.W., Zerihun, M., Messele, A., Richards, F.O., (2008). Integrating an NTD with one of "The Big Three": combined malaria and trachoma survey in Amhara region of Ethiopia. *PLoS Neglected Tropical Diseases* 2 (3), e197.

Etya'ale, D., (2008). Onchocerciasis and trachoma control: what has changed in the past two decades? *Community Eye Health* 21 (67), 43-45.

Faye, M., Kuper, H., Dineen, B., Bailey, R., (2006). Rapid assessment for prioritisation of trachoma control at community level in one district of the Kaolack Region, Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100 (2), 149-157.

Feasey, N., Wansbrough-Jones, M., Mabey, D.C., Solomon, A.W., (2009). Neglected tropical diseases. *British Medical Bulletin* 93 (1), 179-200.

Gabrielli, A.F., Toure, S., Sellin, B., Sellin, E., Ky, C., Ouedraogo, H., Yaogho, M., Wilson, M.D., Thompson, H., Sanou, S., Fenwick, A., (2006). A combined school- and community-based campaign targeting all school-age children of Burkina Faso against schistosomiasis and soil-transmitted helminthiasis: performance, financial costs and implications for sustainability. *Acta Tropica* 99 (2-3), 234-242.

Gemperli, A., (2003). Development of Spatial Statistical Methods for Modelling Point-Referecnced Spatial Data in Malaria Epidemiology. Doctoral dissertation, University of Basel, Basel.

Gething, P.W., Smith, D.L., Patil, A.P., Tatem, A.J., Snow, R.W., Hay, S.I., (2010). Climate change and the global malaria recession. *Nature* 465 (7296), 342-345.

Gething, P.W., Van Boeckel, T.P., Smith, D.L., Guerra, C.A., Patil, A.P., Snow, R.W., Hay, S.I., **(2011).** Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax. Parasit Vectors* 4, 92.

Gitonga, C., Karanja, P., Kihara, J., Mwanje, M., Juma, E., Snow, R., Noor, A., Brooker, S., (2010). Implementing school malaria surveys in Kenya: towards a national surveillance system. *Malaria Journal* 9 (1), 306.

Glinz, D., Silué, K.D., Knopp, S., Lohourignon, L.K., Yao, K.P., Steinmann, P., Rinaldi, L., Cringoli, G., N'Goran, E.K., Utzinger, J., (2010). Comparing diagnostic accuracy of Kato-Katz, Koga agar plate, ether-concentration, and FLOTAC for *Schistosoma mansoni* and soil-transmitted helminths. *PLoS Neglected Tropical Diseases* 4 (7), e754.

Goovaerts, P., (1997). Geostatistics for natural resources evaluation. Oxford University Press, New York.

Gosoniu, L., Vounatsou, P., Sogoba, N., Smith, T., (2006). Bayesian modelling of geostatistical malaria risk data. *Geospatial Health* 1 (1), 127-139.

Gosoniu, L., Vounatsou, P., Sogoba, N., Maire, N., Smith, T., (2009). Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Computational Statistics and Data Analysis* 53 (9), 3358-3371.

Griffin, J.T., Hollingsworth, T.D., Okell, L.C., Churcher, T.S., White, M., Hinsley, W., Bousema, T., Drakeley, C.J., Ferguson, N.M., Basanez, M.-G., Ghani, A.C., (2010). Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies. *PLoS Medicine* 7 (8), e1000324.

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

Guerra, C.A., Gikandi, P.W., Tatem, A.J., Noor, A.M., Smith, D.L., Hay, S.I., Snow, R.W., (2008). The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Medicine* 5 (2), e38.

Gupte, M.D., Murthy, B.N., Mahmood, K., Meeralakshmi, S., Nagaraju, B., Prabhakaran, R., (2004). Application of lot quality assurance sampling for leprosy elimination monitoring-- examination of some critical factors. *International Journal of Epidemiology* 33 (2), 344-348.

Gutman, J., Fagbemi, A., Alphonsus, K., Eigege, A., Miri, E.S., Richards, F.O., Jr., (2008). Missed treatment opportunities for schistosomiasis mansoni, in an active programme for the treatment of urinary schistosomiasis in Plateau and Nasarawa states, Nigeria. *Annals of Tropical Medicine and Parasitology* 102 (4), 335-346.

Gutman, J., Richards, F.O., Jr., Eigege, A., Umaru, J., Alphonsus, K., Miri, E.S., (2009). The presumptive treatment of all school-aged children is the least costly strategy for schistosomiasis control in Plateau and Nasarawa states, Nigeria. *Annals of Tropical Medicine and Parasitology* 103 (6), 501-511.

Guyatt, H.L., Bundy, D.A., Medley, G.F., Grenfell, B.T., (1990). The relationship between the frequency distribution of *Ascaris lumbricoides* and the prevalence and intensity of infection in human communities. *Parasitology* 101 (Pt 1), 139-143.

Guyatt, H.L., Bundy, D.A., (1991). Estimating prevalence of community morbidity due to intestinal helminths: prevalence of infection as an indicator of the prevalence of disease. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85 (6), 778-782.

Gyapong, J.O., Remme, J.H., (2001). The use of grid sampling methodology for rapid assessment of the distribution of bancroftian filariasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95 (6), 681-686.

Gyapong, J.O., Kyelem, D., Kleinschmidt, I., Agbo, K., Ahouandogbo, F., Gaba, J., Owusu-Banahene, G., Sanou, S., Sodahlon, Y.K., Biswas, G., Kale, O.O., Molyneux, D.H., Roungou, J.B., Thomson, M.C., Remme, J., (2002). The use of spatial analysis in mapping the distribution of bancroftian filariasis in four West African countries. *Annals of Tropical Medicine and Parasitology* 96 (7), 695-705.

Hall, A., Hewitt, G., Tuffrey, V., de Silva, N., (2008). A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. *Maternal and Child Nutrition* 4 (Suppl 1), 118-236.

187

Hall, A., Horton, S., de Silva, N., (2009). The costs and cost-effectiveness of mass treatment for intestinal nematode worm infections using different treatment thresholds. *PLoS Negl Trop Dis* 3 (3), e402.

Handzel, T., Karanja, D.M., Addiss, D.G., Hightower, A.W., Rosen, D.H., Colley, D.G., Andove, J., Slutsker, L., Secor, W.E., (2003). Geographic distribution of schistosomiasis and soil-transmitted helminths in Western Kenya: implications for anthelminthic mass treatment. *American Journal of Tropical Medicine and Hygiene* 69 (3), 318-323.

Hay, S.I., Omumbo, J.A., Craig, M.H., Snow, R.W., (2000). Earth observation, geographic information systems and *Plasmodium falciparum* malaria in sub-Saharan Africa. *Advances in Parasitology* 47, 173-215.

Hay, S.I., Graham, A.J., Rogers, D.J., (2006a). Global mapping of infectious diseases: methods, examples and emerging applications. *Advances in Parasitology* 62, ix-xi.

Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J., (2006b). Global environmental data for mapping infectious disease distribution. *Advances in Parasitology* 62, 37-77.

Hay, S.I., Smith, D.L., Snow, R.W., (2008). Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infectious Diseases* 8 (6), 369-378.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W., (2009). A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Medicine* 6 (3), e1000048.

Hay, S.I., Okiro, E.A., Gething, P.W., Patil, A.P., Tatem, A.J., Guerra, C.A., Snow, R.W., (2011). Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PLoS Medicine* 7 (6), e1000290.

Hedt, B.L., Olives, C., Pagano, M., Valadez, J.J., (2008). Large country-Lot Quality Assurance Sampling: a new method for rapid monitoring and evaluation of health, nutrition and population programs at sub-national levels. HNP, World Bank.

Hirzel, A., Guisan, A., (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157 (2-3), 331-341.

Homeida, M., Zimmerman, V., Baraka, O.Z., Sulieman, S.M., Awad, K., Omer, K., Ali, M.M.A., Mustafa, B., (1994). Helminth infections in Sudan. *The Lancet* 344, 1228.

Hooper, P.J., Bradley, M.H., Biswas, G., Ottesen, E.A., (2009). The Global Programme to Eliminate Lymphatic Filariasis: health impact during its first 8 years (2000-2007). *Annals of Tropical Medicine and Parasitology* 103 Suppl 1, S17-21.

Hopkins, D., Eigege, A., Miri, E.S., Gontor, I., Ogah, G., Umaru, J., Gwomkudu, C.C., Mathai, W., Jinadu, M., Amadiegwu, S., Oyenekan, O.K., Korve, K., Richards, F.O., Jr., (2002). Lymphatic filariasis elimination and schistosomiasis control in combination with onchocerciasis control in Nigeria. *American Journal of Tropical Medicine and Hygiene* 67 (3), 226-272.

Hotez, P.J., Bundy, D.A.P., Beegle, K., Brooker, S., Drake, L., de Silva, N., Montresor, A., Engels, D., Jukes, M., Chitsulo, L., Chow, J., Laxminarayan, R., Michaud, C., Bethony, J., Oliveira, R., Xiao, S.H., Fenwick, A., Savioli, L., (2006a). Helminth infections: soil-transmitted helminth infections and schistosomiasis. In: *Disease Control Priorities in Developing Countries* edited by Jamison, D.T., Evans, D.B., Alleyne, G., Jha, P., Breman, J., Measham, A.R., Mariam, C., Mills, A.,Musgrove, P.R., Oxford University Press, New York. pp. 467-497.

Hotez, P.J., Molyneux, D.H., Fenwick, A., Ottesen, E., Ehrlich Sachs, S., Sachs, J.D., (2006b). Incorporating a rapid-impact package for Neglected Tropical Diseases with programs for HIV/AIDS, tuberculosis, and malaria. *PLoS Medicine* 3 (5), e102.

Howard, S.C., Donnell, C.A., Chan, M.S., (2001). Methods for estimation of associations between multiple species parasite infections. *Parasitology* 122 (Pt 2), 233-251.

Hutin, Y.J., Legros, D., Owini, V., Brown, V., Lee, E., Mbulamberi, D., Paquet, C., (2004). *Trypanosoma brucei gambiense* trypanosomiasis in Terego county, northern Uganda, 1996: a lot quality assurance sampling survey. *American Journal of Tropical Medicine and Hygiene* 70 (4), 390-394.

Hviid, L., Staalsoe, T., (2004). Malaria immunity in infants: a special case of a general phenomenon? *Trends in Parasitology* 20 (2), 66-72.

Jordan, P., Webbe, G., (1993). Epidemiology. In: *Human schistosomiasis* edited by Jordan, P., Webbe, G.,Sturrock, R.F., CAB International, Wallingford. pp. 87-158.

Jovani, R., Tella, J.L., (2006). Parasite prevalence and sample size: misconceptions and solutions. *Trends in Parasitology* 22 (5), 214-218.

Kabatereine, N.B., Brooker, S., Tukahebwa, E.M., Kazibwe, F., Onapa, A.W., (2004). Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. *Tropical Medicine and International Health* 9 (3), 372-380.

Kabatereine, N.B., Tukahebwa, E.M., Kazibwe, F., Twa-Twa, J.M., Barenzi, J.F., Zaramba, S., Stothard, J.R., Fenwick, A., Brooker, S., (2005). Soil-transmitted helminthiasis in Uganda: epidemiology and cost of control. *Tropical Medicine and International Health* 10 (11), 1187-1189.

Kazembe, L.N., Kleinschmidt, I., Holtz, T.H., Sharp, B.L., (2006). Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *International Journal of Health Geographics* 5, 41.

Kazura, J.W., (2002). Lymphatic filarial infections: an introduction to the filariae. In: *The filaria* edited by Klei, T.R.,Rajan, T.V., Kluwer Academic Publishers.

King, J.D., Ngondi, J., Gatpan, G., Lopidia, B., Becknell, S., Emerson, P.M., (2008). The burden of trachoma in Ayod County of Southern Sudan. *PLoS Neglected Tropical Diseases* 2 (9), e299.

King, J.D., Eigege, A., Richards, F., Jr., Jip, N., Umaru, J., Deming, M., Miri, E., McFarland, D., Emerson, P.M., (2009). Integrating NTD mapping protocols: can surveys for trachoma and urinary schistosomiasis be done simultaneously? *American Journal of Tropical Medicine and Hygiene* 81 (5), 793-798.

Kleinschmidt, I., Bagayoko, M., Clarke, G.P., Craig, M., Le Sueur, D., (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* 29 (2), 355-361.

Kleinschmidt, I., Omumbo, J., Briet, O., van de Giesen, N., Sogoba, N., Mensah, N., Windmeijer, P., Moussa, M., Teuscher, T., (2001). An empirical malaria distribution map for West Africa. *Tropical Medicine and International Health* 6, 779-786.

Kloos, H., Lo, C.T., Birrie, H., Ayele, T., Tedla, S., Tsegay, F., (1988). Schistosomiasis in Ethiopia. *Social Science and Medicine* 26 (8), 803-827.

Kloos, H., Gazzinelli, A., Van Zuyle, P., (1998). Microgeographical patterns of schistosomiasis and water contact behavior; examples from Africa and Brazil. *Memórias do Instituto Oswaldo Cruz* 93 Suppl 1, 37-50.

Knopp, S., Mgeni, A.F., Khamis, I.S., Steinmann, P., Stothard, J.R., Rollinson, D., Marti, H., Utzinger, J., (2008). Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. *PLoS Neglected Tropical Diseases* 2 (11), e331.

Knopp, S., Rinaldi, L., Khamis, I.S., Stothard, J.R., Rollinson, D., Maurelli, M.P., Steinmann, P., Marti, H., Cringoli, G., Utzinger, J., (2009). A single FLOTAC is more sensitive than triplicate Kato-Katz for the diagnosis of low-intensity soil-transmitted helminth infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103 (4), 347-354.

Kolaczinski, J.H., Kolaczinski, K., Kyabayinze, D., Strachan, D., Temperley, M., Wijayanandana, N., Kilian, A., (2010). Costs and effects of two public sector delivery channels for long-lasting insecticidal nets in Uganda. *Malaria Journal* 9, 102.

Kur, L.W., Picon, D., Adibo, O., Robinson, E., Sabasio, A., Edwards, T., Ndyaba, A., Rumunu, J., Lewis, K., Lado, M., Kolaczinski, J., (2009). Trachoma in Western Equatoria State, Southern Sudan: implications for national control. *PLoS Neglected Tropical Diseases* 3 (7), e492.

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.

Lanata, C.F., Stroh, G., Jr., Black, R.E., Gonzales, H., (1990). An evaluation of lot quality assurance sampling to monitor and improve immunization coverage. *International Journal of Epidemiology* 19 (4), 1086-1090.

Lark, R.M., (2000). Designing sampling grids from imprecise information on soil variability, an approach based on the fuzzy kriging variance. *Geoderma* 98, 35-59.

Legendre, P., (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74, 1659–1673.

Legendre, P., Dale, M.R.T., Fortin, M.J., Gurevitch, J., Hohn, M., Myers, D., (2002). The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25 (5), 601-615.

Legesse, M., Erko, B., (2007). Field-based evaluation of a reagent strip test for diagnosis of *Schistosoma mansoni* by detecting circulating cathodic antigen in urine before and after chemotherapy. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 101 (7), 668-673.

Lemeshow, S., Taber, S., (1991). Lot quality assurance sampling: single- and double-sampling plans. *World Health Statistics Quarterly. Rapport Trimestriel de Statistiques Sanitaires Mondiales* 44 (3), 115-132.

Lengeler, C., Kilima, P., Mshinda, H., Morona, D., Hatz, C., Tanner, M., (1991). Rapid, low-cost, two-step method to screen for urinary schistosomiasis at the district level: the Kilosa experience. *Bulletin of the World Health Organization* 69 (2), 179-189.

Lengeler, C., Utzinger, J., Tanner, M., (2002). Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa. *Bulletin of the World Health Organization* 80 (3), 235-242.

Levy, P.S., Lemeshow, S., (1999). Sampling of populations: methods and applications. Wiley.

Liese, B., Rosenberg, M., Schratz, A., (2010). Programmes, partnerships, and governance for elimination and control of neglected tropical diseases. *Lancet* 375 (9708), 67-76.

Lindsay, S.W., Thomas, C.J., (2000). Mapping and estimating the population at risk from lymphatic filariasis in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 94 (1), 37-45.

Lloyd, C.D., Atkinson, P.M., 1998. Scale and the spatial structure of landform: optimising sampling strategies with geostatistics. In, Proceedings of the 3rd International Conference on GeoComputation, http://www.geocomputation.org/1998/15/gc 15.htm.

Lwambo, N.J.S., Siza, J.E., Brooker, S., Bundy, D.A.P., Guyatt, H., (1999). Patterns of concurrent infection with hookworm and schistosomiasis in school children in Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 93, 497-502.

Lwanga, S.K., Lemeshow, S., (1991). Sample size determination in health studies. WHO, Geneva.

Mabaso, M.L., Appleton, C.C., Hughes, J.C., Gouws, E., (2003). The effect of soil type and climate on hookworm (*Necator americanus*) distribution in KwaZulu-Natal, South Africa. *Tropical Medicine and International Health* 8 (8), 722-727.

Mabey, D.C.W., Solomon, A.W., Foster, A., (2003). Trachoma. *The Lancet* 362 (9379), 223-229.

Magalhães, R.J.S., Clements, A.C.A., Patil, A.P., Gething, P.W., Brooker, S., (2011). The applications of model-based geostatistics in helminth epidemiology and control. *Advances in Parasitology* 74, 267-296.

Magambo, J.K., Zeyhle, E., Wachira, T.M., (1998). Prevalence of intestinal parasites among children in southern Sudan. *East African Medical Journal* 75 (5), 288-290.

Mathieu, E., Deming, M., Lammie, P.J., McLaughlin, S.I., Beach, M.J., Domingue, J., Addiss, D.G., (2003). Comparison of methods for estimating drug coverage for filariasis elimination, Leogane Commune, Haiti. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 97 (5), 501-505.

Mbogo, C.M., Mwangangi, J.M., Nzovu, J., Gu, W., Yan, G., Gunter, J.T., Swalm, C., Keating, J., Regens, J.L., Shililu, J.I., Githure, J.I., Beier, J.C., (2003). Spatial and temporal

heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenyan coast. *American Journal of Tropical Medicine and Hygiene* 68 (6), 734-742.

Mbogo, C.N., Snow, R.W., Khamala, C.P., Kabiru, E.W., Ouma, J.H., Githure, J.I., Marsh, K., Beier, J.C., (1995). Relationships between *Plasmodium falciparum* transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *American Journal of Tropical Medicine and Hygiene* 52 (3), 201-206.

Mecaskey, J.W., Knirsch, C.A., Kumaresan, J.A., Cook, J.A., (2003). The possibility of eliminating blinding trachoma. *Lancet Infectious Diseases* 3 (11), 728-734.

Michael, E., Grenfell, B.T., Bundy, D.A., (1994). The association between microfilaraemia and disease in lymphatic filariasis. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 256 (1345), 33-40.

Michael, E., Bundy, D.A.P., (1997). Global mapping of lymphatic filariasis. *Parasitology Today* 13 (12), 472-476.

Michael, E., (2000). The population dynamics and epidemiology of lymphatic filariasis. In: *Lymphatic filariasis* edited by Nutman, T., Imperial College Press, London.

Midzi, N., Butterworth, A.E., Mduluza, T., Munyati, S., Deelder, A.M., van Dam, G.J., (2009). Use of circulating cathodic antigen strips for the diagnosis of urinary schistosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103 (1), 45-51.

MoH-GOSS, (2009). National Survey on Neglected Tropical Diseases in Southern Sudan. Survey Protocol. Ministry of Health, Government of Southern Sudan, Revised February 2009. *http://www.malariaconsortium.org/userfiles/NTD%20Mapping%20S%20Sudan%20Feb%2009%20Revised.pdf*.

Molyneux, D.H., Hotez, P.J., Fenwick, A., (2005). "Rapid-impact interventions": how a policy of integrated control for Africa's neglected tropical diseases could benefit the poor. *PLoS Medicine* 2 (11), e336.

Molyneux, D.H., Hotez, P.J., Fenwick, A., Newman, R.D., Greenwood, B., Sachs, J., (2009). Neglected tropical diseases and the Global Fund. *Lancet* 373 (9660), 296-297.

Montresor, A., Crompton, D.W.T., Hall, A., Bundy, D.A.P., Savioli, L., (1998). Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level. World Health Organisation, Geneva.

Muennig, P., Pallin, D., Sell, R.L., Chan, M.S., (1999). The cost effectiveness of strategies for the treatment of intestinal parasites in immigrants. *New England Journal of Medicine* 340 (10), 773-779.

Murray, C.K., Gasser, R.A., Jr., Magill, A.J., Miller, R.S., (2008). Update on rapid diagnostic testing for malaria. *Clinical Microbiology Reviews* 21 (1), 97-110.

Mwangi, T.W., Bethony, J.M., Brooker, S., (2006). Malaria and helminth interactions in humans: an epidemiological viewpoint. *Annals of Tropical Medicine and Parasitology* 100 (7), 551-570.

Myatt, M., Limburg, H., Minassian, D., Katyola, D., (2003). Field trial of applicability of lot quality assurance sampling survey method for rapid assessment of prevalence of active trachoma. *Bulletin of the World Health Organisation* 81 (12), 877-885.

Myatt, M., Mai, N.P., Quynh, N.Q., Nga, N.H., Tai, H.H., Long, N.H., Minh, T.H., Limburg, H., (2005). Using lot quality-assurance sampling and area sampling to identify priority areas for trachoma control: Viet Nam. *Bulletin of the World Health Organisation* 83 (10), 756-763.

Na-Bangchang, K., Kietinun, S., Pawa, K.K., Hanpitakpong, W., Na-Bangchang, C., Lazdins, J., (2006). Assessments of pharmacokinetic drug interactions and tolerability of albendazole, praziquantel and ivermectin combinations. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 100 (4), 335-345.

Negrel, A.D., Taylor, H.R., West, S., (2001). Guidelines for rapid assessment for blinding trachoma. WHO, Geneva.

Ngondi, J., Ole-Sempele, F., Onsarigo, A., Matende, I., Baba, S., Reacher, M., Matthews, F., Brayne, C., Emerson, P.M., (2006). Prevalence and causes of blindness and low vision in southern Sudan. *PLoS Medicine* 3 (12), e477.

Noma, M., Nwoke, B.E., Nutall, I., Tambala, P.A., Enyong, P., Namsenmo, A., Remme, J., Amazigo, U.V., Kale, O.O., Seketeli, A., (2002). Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). *Annals of Tropical Medicine and Parasitology* 96 (Suppl 1), S29-39.

O'Meara, W.P., Mangeni, J.N., Steketee, R., Greenwood, B., (2010). Changes in the burden of malaria in sub-Saharan Africa. *Lancet Infectious Diseases* 10 (8), 545-555.

Okell, L.C., Ghani, A.C., Lyons, E., Drakeley, C.J., (2009). Submicroscopic Infection in *Plasmodium falciparum*–Endemic Populations: A Systematic Review and Meta Analysis. *The Journal of Infectious Diseases* 200 (10), 1509-1517.

Oliver, M.A., Muir, K.R., Webster, R., Parkes, S.E., Cameron, A.H., Stevens, M.C., Mann, J.R., (1992). A geostatistical approach to the analysis of pattern in rare disease. *Journal of Public Health Medicine* 14 (3), 280-289.

Olives, C., Pagano, M., Deitchler, M., Hedt, B.L., Egge, K., Valadez, J.J., (2009). Cluster designs to assess the prevalence of acute malnutrition by lot quality assurance sampling: a validation study by computer simulation. *Journal of the Royal Statistical Society (Series C): Applied Statistics* 172, 495-510.

Olsen, A., (2007). Efficacy and safety of drug combinations in the treatment of schistosomiasis, soil-transmitted helminthiasis, lymphatic filariasis and onchocerciasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 101 (8), 747-758.

Olson, D.M., Dinerstein, E., (2009). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden* 89 (2), 125-126.

Ongom, V.L., Bradley, D.J., (1972). The epidemiology and consequences of *Schistosoma mansoni* infection in West Nile, Uganda. I. Field studies of a community at Panyagoro. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 66 (6), 835-851.

Pfeiffer, D.U., Robinson, T.P., Stevenson, M., Stevens, K.B., Rogers, D.J., Clements, A.C.A., (2008). Spatial analysis in epidemiology. OUP, New York.

Polack, S., Brooker, S., Kuper, H., Mariotti, S., Mabey, D., Foster, A., (2005). Mapping the global distribution of trachoma. *Bulletin of the World Health Organization* 83 (12), 913-919.

R Development Core Team, (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabarijaona, L., Rakotomanana, F., Ranaivo, L., Raharimalala, L., Modiano, D., Boisier, P., De Giorgi, F., Raveloson, N., Jambou, R., (2001). Validity of lot quality assurance sampling to optimize *falciparum* malaria surveys in low-transmission areas. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95 (3), 267-269.

Raso, G., Matthys, B., N'Goran E, K., Tanner, H.P., Vounatsou, P., Utzinger, J., (2005). Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Cote d'Ivoire. *Parasitology* 131 (Pt 1), 97-108.

Raso, G., Vounatsou, P., Gosoniu, L., Tanner, M., N'Goran, E.K., Utzinger, J., (2006a). Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Cote d'Ivoire. *International Journal for Parasitology* 36 (2), 201-210.

Raso, G., Vounatsou, P., Singer, B.H., N'Goran, E.K., Tanner, M., Utzinger, J., (2006b). An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. *Proceedings of the National Academy of Sciences of the United States of America* 103 (18), 6934-6939.

Raso, G., Silue, K.D., Vounatsou, P., Singer, B.H., Yapi, A., Tanner, M., Utzinger, J., N'Goran, E.K., (2009). Spatial risk profiling of *Plasmodium falciparum* parasitaemia in a high endemicity area in Cote d'Ivoire. *Malaria Journal* 8, 252.

Ratard, R.C., Kouemeni, L.E., Bessala, M.M.E., Ndamkou, C.N., Sama, M.T., Cline, B.L., (1991). Ascariasis and trichuriasis in Cameroon. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85 (1), 84-88.

Ratard, R.C., Kouemeni, L.E., Ekani Bessala, M.K., Ndamkou, C.N., (1992). Distribution of hookworm infection in Cameroon. *Annals of Tropical Medicine and Parasitology* 86 (4), 413-418.

Ribeiro Jnr, P.J., Diggle, P.J., (2001). geoR: A package for geostatistical analysis. *R-NEWS* 1 (2), 15-18.

Robertson, S.E., Valadez, J.J., (2006). Global review of health care surveys using lot quality assurance sampling (LQAS), 1984-2004. *Social Science and Medicine* 63 (6), 1648-1660.

Robinson, E., Picon, D., Sturrock, H.J.W., Sabasio, A., Lado, M., Kolaczinski, J.H., Brooker, S., (2009). The performance of haematuria reagent strips for the rapid mapping of urinary schistosomiasis: field experience from Southern Sudan. *Tropical Medicine & International Health* 14 (12), 1484-1487.

Robinson, T.P., (1998). Geographic information systems and the selection of priority areas for control of tsetse-transmitted trypanosomiasis in Africa. *Parasitology Today* 14 (11), 457-461.

Rogers, D.J., Randolph, S.E., (1993). Distribution of tsetse and ticks in Africa: past, present and future. *Parasitology Today* 9 (7), 266-271.

Roll Back Malaria Monitoring and Evaluation Reference Group, (2005). Malaria indicator survey: basic documentation for survey design and implementation. Geneva.

Ross, A.G., Bartley, P.B., Sleigh, A.C., Olds, G.R., Li, Y., Williams, G.M., McManus, D.P., (2002). Schistosomiasis. *New England Journal of Medicine* 346 (16), 1212-1220.

Rumunu, J., Brooker, S., Hopkins, A., Chane, F., Emerson, P., Kolaczinski, J., (2009). Southern Sudan: an opportunity for NTD control and elimination? *Trends in Parasitology* 25 (7), 301-307.

Saathoff, E., Olsen, A., Sharp, B., Kvalsvig, J.D., Appleton, C.C., Kleinschmidt, I., (2005). Ecologic covariates of hookworm infection and reinfection in rural Kwazulu-natal/south Africa: a

geographic information system-based study. *American Journal of Tropical Medicine and Hygiene* 72 (4), 384-391.

Schad, G.A., Anderson, R.M., (1985). Predisposition to hookworm infection in humans. *Science* 228 (4707), 1537-1540.

Shope, R.E., (1999). Factors influencing the geographic distribution and incidence of tropical infectious diseases. In: *Tropical infectious diseases. Principles, pathogens and practice* edited by Guerrant, R.L., Walker, D.H.,Weller, P.F. 1, Churchill Livingstone, London. pp. 16-21.

Simoonga, C., Utzinger, J., Brooker, S., Vounatsou, P., Appleton, C.C., Stensgaard, A.S., Olsen, A., Kristensen, T.K., (2009). Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology* 136 (13), 1683-1693.

Singh, J., Jain, D.C., Sharma, R.S., Verghese, T., (1996). Evaluation of immunization coverage by lot quality assurance sampling compared with 30-cluster sampling in a primary health centre in India. *Bulletin of the World Health Organisation* 74 (3), 269-274.

Smith, D.L., Dushoff, J., Snow, R.W., Hay, S.I., (2005). The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* 438 (7067), 492-495.

Smith, D.L., Guerra, C.A., Snow, R.W., Hay, S.I., (2007a). Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malaria Journal* 6 (1), 131.

Smith, D.L., McKenzie, F.E., Snow, R.W., Hay, S.I., (2007b). Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol* 5 (3), e42.

Smith, D.L., Drakeley, C.J., Chiyaka, C., Hay, S.I., (2010). A quantitative analysis of transmission efficiency versus intensity for malaria. *Nature Communications* 1, 108.

Smith, J.L., Brooker, S., (2010). Impact of hookworm infection and deworming on anaemia in non-pregnant populations: a systematic review. *Tropical Medicine and International Health* 15 (7), 776-795.

Snow, R.W., Marsh, K., le Sueur, D., (1996). The need for maps of transmission intensity to guide malaria control in Africa. *Parasitology Today* 12, 455-457.

Snow, R.W., Omumbo, J.A., (2006). Malaria. In: *Disease and Mortality in Sub-Saharan Africa* (2nd), edited by Jamison, D., Feacham, R., Makgoba, M., Bos, E., Baingana, F., Hofman, K.,Rogo, K., The World Bank, Washington D.C.

Snow, R.W., Marsh, K., (2010). Malaria in Africa: progress and prospects in the decade since the Abuja Declaration. *Lancet* 376 (9735), 137-139.

Srividya, A., Michael, E., Palaniyandi, M., Pani, S.P., Das, P.K., (2002). A geostatistical analysis of the geographic distribution of lymphatic filariasis prevalence in southern India. *American Journal of Tropical Medicine and Hygiene* 67 (5), 480-489.

Standley, C.J., Lwambo, N.J.S., Lange, C.N., Kariuki, H.C., Adriko, M., Stothard, J.R., (2010). Performance of circulating cathodic antigen (CCA) urine-dipsticks for rapid detection of intestinal schistosomiasis in schoolchildren from shoreline communities of Lake Victoria. *Parasites and Vectors* 3 (1), 7.

Stein, A., Ettema, C., (2003). An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. *Agriculture, Ecosystems and Environment* 94 (1), 31-47.

Stephenson, L.S., Latham, M.C., Ottesen, E.A., (2000). Malnutrition and parasitic helminth infections. *Parasitology* 121 (Suppl), S23-38.

Stothard, J.R., Kabatereine, N.B., Tukahebwa, E.M., Kazibwe, F., Rollinson, D., Mathieson, W., Webster, J.P., Fenwick, A., (2006). Use of circulating cathodic antigen (CCA) dipsticks for detection of intestinal and urinary schistosomiasis. *Acta Tropica* 97 (2), 219-228.

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.

Stothard, J.R., Sousa-Figueiredo, J.C., Standley, C.J., Van Dam, G.J., Knopp, S., Utzinger, J., Ameri, H., Khamis, A.N., Khamis, I.S., Deelder, A.M., Mohammed, K.A., Rollinson, D., (2009). An evaluation of urine-CCA strip test and fingerprick blood SEA-ELISA for detection of urinary schistosomiasis in schoolchildren in Zanzibar. *Acta Tropica* 111 (1), 64-70.

Sturrock, H.J., Picon, D., Sabasio, A., Oguttu, D., Robinson, E., Lado, M., Rumunu, J., Brooker, S., Kolaczinski, J.H., (2009). Integrated mapping of neglected tropical diseases: epidemiological findings and control implications for northern Bahr-el-Ghazal State, Southern Sudan. *PLoS Neglected Tropical Diseases* 3 (10), e537.

Sturrock, H.J.W., Gething, P.W., Clements, A.C., Brooker, S., (2010). Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-efficiency of sampling. *American Journal of Tropical Medicine and Hygiene* 82 (6), 1079-1087.

Sturrock, R.F., (1993). The Intermediate Hosts and Host- Parasite Relationships. In: *Human schistosomiasis* edited by Jordan, P., Webbe, G.,Sturrock, R.F., CAB International, Wallingford. pp. 33-85.

Takougang, I., Meremikwu, M., Wandji, S., Yenshu, E.V., Aripko, B., Lamlenn, S.B., Eka, B.L., Enyong, P., Meli, J., Kale, O., Remme, J.H., (2002). Rapid assessment method for prevalence and intensity of *Loa loa* infection. *Bulletin of the World Health Organisation* 80 (11), 852-858.

Taylor-Robinson, D.C., Jones, A.P., Garner, P., (2007). Deworming drugs for treating soil-transmitted intestinal worms in children: effects on growth and school performance. *Cochrane Database Systematic Reviews* (4), CD000371.

TDR, (2002). Guidelines for rapid assessment of Loa loa. UNDP/World Bank/WHO Special Programme for Research & Training in Tropical Diseases (TDR), Geneva

Temperley, M., Mueller, D.H., Njagi, J.K., Akhwale, W., Clarke, S.E., Jukes, M.C., Estambale, B.B., Brooker, S., (2008). Costs and cost-effectiveness of delivering intermittent preventive treatment through schools in western Kenya. *Malaria Journal* 7, 196.

Thomson, M.C., Connor, S.J., Dalessandro, U., Rowlingson, B., Diggle, P., Cresswell, M., Greenwood, B., (1999). Predicting malaria infection in Gambian children from satellite data and bed net use surveys: The importance of spatial correlation in the interpretation of results. *American Journal of Tropical Medicine and Hygiene* 61 (1), 2-8.

Thomson, M.C., Obsomer, V., Kamgno, J., Gardon, J., Wanji, S., Takougang, I., Enyong, P., Remme, J.H., Molyneux, D.H., Boussinesq, M., (2004). Mapping the distribution of *Loa loa* in Cameroon in support of the African Programme for Onchocerciasis Control. *Filaria Journal* 3 (1), 7.

Tobler, W.R., (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234-240.

Tun-Lin, W., Burkot, T.R., Kay, B.H., (2000). Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes aegypti* in north Queensland, Australia. *Medical and Veterinary Entomology* 14 (1), 31-37.

Udonsi, J.K., Nwosu, A.B.C., Anya, A.O., (1980). *Necator americanus*: population structure, distribution, and fluctuations in population densities of infective larvae in contaminated farmlands. *Zeitschrift fur Parasitenkunde* 63, 251-259.

Utzinger, J., Muller, I., Vounatsou, P., Singer, B.H., N'Goran, E.K., Tanner, M., (2003). Random spatial distribution of *Schistosoma mansoni* and hookworm infections among school children within a single village. *Journal of Parasitology* 89 (4), 686-692.

Utzinger, J., Rinaldi, L., Lohourignon, L.K., Rohner, F., Zimmermann, M.B., Tschannen, A.B., N'Goran, E.K., Cringoli, G., (2008). FLOTAC: a new sensitive technique for the diagnosis of hookworm infections in humans. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 102 (1), 84-90.

Van Groenigen, J.W., Siderius, W., Stein, A., (1998). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87, 239-259.

Van Groenigen, J.W., Stein, A., (1999). Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* 27 (5), 1078-1086.

Van Groenigen, J.W., (2000). The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma* 97, 223-236.

Vlassoff, C., Tanner, M., (1992). The relevance of rapid assessment to health research and interventions. *Health Policy and Planning* 7 (1), 1-9.

Walker, D., Kumaranayake, L., (2002). Allowing for differential timing in cost analyses: discounting and annualization. *Health Policy and Planning* 17 (1), 112-118.

Warren, K.S., Bundy, D.A.P., Anderson, R.M., Davis, A.R., Henderson, D.A., Jamison, D.T., Prescott, N., Senft, A., (1993). Helminth infections. In: *In Disease Control Priorities in Developing Countries* edited by Jamison, D.T., Mosley, W.H., Measham, A.R.,Bobadilla, J.L., Oxford University Press, Oxford. pp. 131-160.

WHO, (1987). Atlas of the global distribution of schistosomiasis. Available from: http://www.who.int/wormcontrol/documents/maps/en/. WHO, Geneva

WHO, (1991). Basic laboratory methods in medical parasitology. World Health Organization, Geneva.

WHO, (1998). Research on rapid geographical assessment Bancroftian filariasis. Available from: http://apps.who.int/tdr/publications/tdr-research-publications/bancroftian-filariasis/pdf/bancroftian-filariasis.pdf.

WHO, (2000a). Operational guidelines for rapid mapping of Bancroftian filariasis in Africa. World Health Organisation, Geneva.

WHO, (2000b). Preparing and implementing a national plan to eliminate lymphatic filariasis. World Health Organization, Geneva.

WHO, (2002). Reducing risks, promoting healthy life. The World Health Report 2002. *Annex Table 3. Burden of disease in DALYs by cause, sex and mortality stratums in WHO regions, estimates for 2001*, 192.

WHO, (2005). Monitoring and epidemiological assessment of the programme to eliminate lymphatic filariasis at implementation unit level. World Health Organization, Geneva.

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

WHO, (2006b). Trachoma control: A guide for programme managers. World Health Organization, Geneva.

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

WHO, (2008). The Global Burden of Disease: 2004 update. World Health Organizaton, Geneva.

WHO, (2009a). Malaria rapid diagnostic test performance: results of WHO product testing of malaria RDTs: round 1 (2008). http://www.finddiagnostics.org/export/sites/default/media/press/pdf/Full-report-malaria-RDTs.pdf. World Health Organization, Geneva

WHO, (2009b). Malaria rapid diagnostic test performance: results of WHO product testing of malaria RDTs: round 2 (2009). http://www.finddiagnostics.org/export/sites/default/resource-centre/reports brochures/docs/rdt malariaRd2 web.pdf. World Health Organization, Geneva

Williams, B., Gopi, P.G., Borgdorff, M.W., Yamada, N., Dye, C., (2008). The design effect and cluster samples: optimising tuberculosis prevalence surveys. International Journal of Tuberculosis and Lung Disease 12 (10), 1110-1115.

World Bank, (1993). World Development Report 1993: Investing in Health. Oxford University Press, New York.

World Bank, (2004). Trachoma at a glance. http://qo.worldbank.org/0Q3WMQ2KQ0.

Zhou, G., Githeko, A., Minakawa, N., Yan, G., (2010). Community-wide benefits of targeted indoor residual spray for malaria control in the Western Kenya Highland. *Malaria Journal* 9 (1), 67.