International Health 3 (2011) 165–175

Contents lists available at ScienceDirect

International Health

journal homepage: http://www.elsevier.com/locate/inhe

Planning schistosomiasis control: investigation of alternative sampling strategies for *Schistosoma mansoni* to target mass drug administration of praziquantel in East Africa

Hugh J.W. Sturrock[a,*], Pete W. Gething[b], Ruth A. Ashton[c,d], Jan H. Kolaczinski[a,d], Narcis B. Kabatereine[e], Simon Brooker[a,f]

[a] *Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom*
[b] *Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom*
[c] *Malaria Consortium – Ethiopia Office, Addis Ababa, Ethiopia*
[d] *Malaria Consortium – Africa Regional Office, Kampala, Uganda*
[e] *National Schistosomiasis and Soil-Transmitted Helminthiasis Control Programme, Vector Control Division, Ministry of Health, Kampala, Uganda*
[f] *Malaria Public Health & Epidemiology Group, Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya*

ARTICLE INFO

Article history:
Received 21 September 2010
Received in revised form 28 February 2011
Accepted 15 June 2011

Keywords:
Schistosoma mansoni
School surveys
Sampling
Lot quality assurance sampling
Geostatistics
Kenya
Ethiopia

ABSTRACT

In schistosomiasis control, there is a need to geographically target treatment to populations at high risk of morbidity. This paper evaluates alternative sampling strategies for surveys of *Schistosoma mansoni* to target mass drug administration in Kenya and Ethiopia. Two main designs are considered: lot quality assurance sampling (LQAS) of children from all schools; and a geostatistical design that samples a subset of schools and uses semi-variogram analysis and spatial interpolation to predict prevalence in the remaining unsurveyed schools. Computerized simulations are used to investigate the performance of sampling strategies in correctly classifying schools according to treatment needs and their cost-effectiveness in identifying high prevalence schools. LQAS performs better than geostatistical sampling in correctly classifying schools, but at a cost with a higher cost per high prevalence school correctly classified. It is suggested that the optimal surveying strategy for *S. mansoni* needs to take into account the goals of the control programme and the financial and drug resources available.

1. Introduction

Control strategies against human schistosomiasis, which focus predominantly on mass drug administration (MDA) of the anthelmintic praziquantel, are most cost-effective when targeted to communities with the highest prevalence of infection and presumed greatest morbidity.[1,2] The World Health Organization (WHO) currently recommends mass treatment of all school-age children once every two years in areas where prevalence exceeds 10%.[3] For *Schistosoma haematobium*, geographical targeting of treatment can be effectively and rapidly achieved through questionnaire-based studies administered by teachers to school children[1] generating data on the presence of blood in urine (a well-established marker of infection). Questionnaires based on reported blood in stools are less reliable for intestinal schistosomiasis caused by *S. mansoni*. Therefore, parasitological examination of stool samples remains the recommended diagnostic method, but this method is time-consuming and expensive.[2] Lot quality assurance sampling (LQAS), which uses small sample sizes to classify communities

* Corresponding author.
E-mail address: hugh.sturrock@lshtm.ac.uk (H.J.W. Sturrock).

1876-3413/$ – see front matter © 2011 Royal Society of Tropical Medicine and Hygiene. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.inhe.2011.06.002

according to prevalence, is one approach to minimising the time and resources needed to conduct parasitological surveys and has been shown to be more cost-effective than the delivery of blanket MDA without prior surveys.[4] 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 paper, we investigate two sampling designs for *S. mansoni* surveys that aim to identify schools requiring MDA in known endemic regions. Specifically, we compare LQAS to a geostatistical survey design which collects data on a subset of schools and uses this information to predict prevalence at unsurveyed schools. We also evaluate the ecological limits of parasite transmission and so reduce the size of the sampling frame within which the two survey designs are implemented. Finally, we incorporate cost estimates of both surveys and subsequent MDA campaigns implemented on the basis of survey results to estimate the cost-effectiveness of the alternative designs.

2. Methods

2.1. Study settings

This analysis focuses on Oromia Regional State in Ethiopia and Western and Nyanza provinces in Kenya (Figure 1). These areas were chosen because of: the widespread occurrence of *S. mansoni*; the availability of geo-referenced prevalence data on *S. mansoni*; and the existence of geo-referenced databases of all government primary schools. In Oromia, there are 5251 government primary schools, and in Western and Nyanza provinces there are a total of 5695 government primary schools.

2.2. Simulation of a 'gold standard' data set

In order to generate a gold standard pseudo-dataset with realistic spatial and aspatial characteristics, we first used data from across Kenya and Ethiopia, derived from the Global Atlas of Helminth Infection,[5,6] 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 primary schools

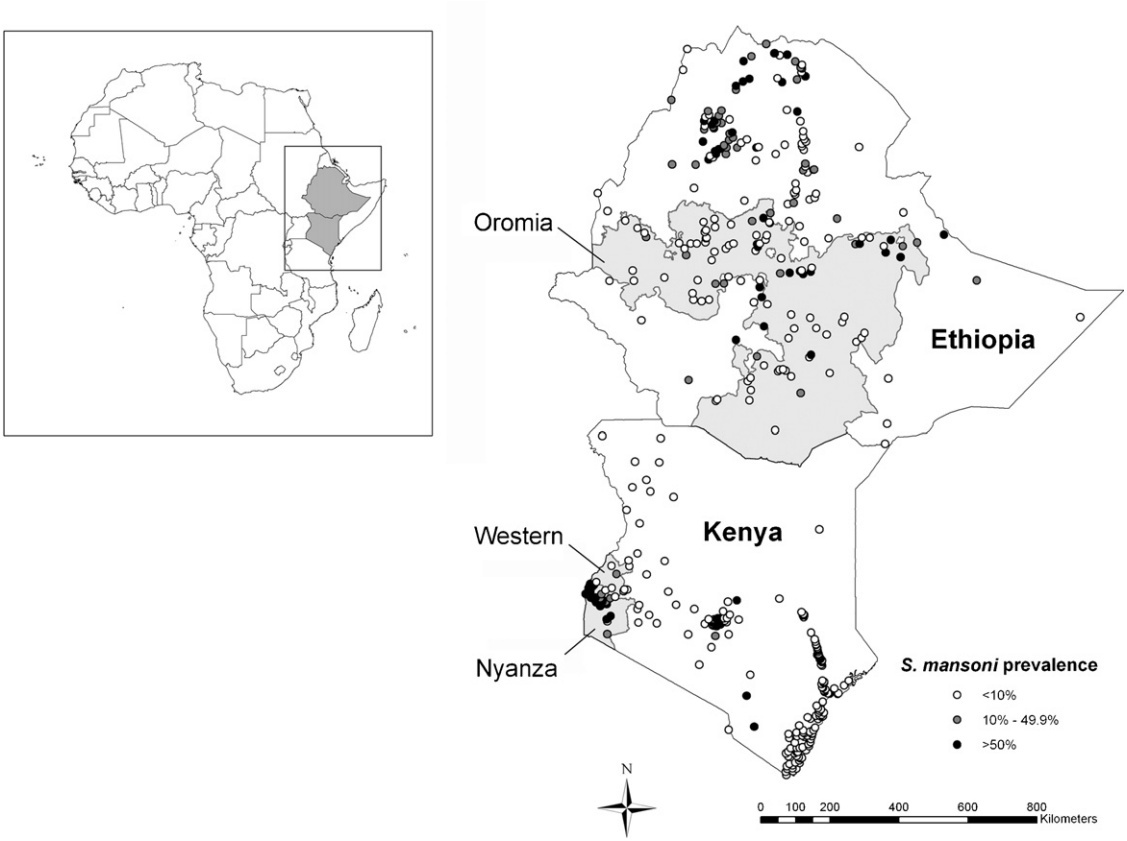

Figure 1. Map of surveyed primary schools where Kato-Katz was used in Kenya ($n = 385$) and Ethiopia ($n = 215$) included in the present analysis. Data were derived from a Global Atlas of Helminth Infection.[5,6] The shaded regions in each country indicate the provinces considered in this study. Inset map: positions of Ethiopia and Kenya within Africa.

Table 1
Thresholds of the environmental variables used to describe the limits of *Schistosoma mansoni* transmission in Western Kenya and Ethiopia

Variable	Western Kenya		Ethiopia	
	Lower limit	Upper limit	Lower limit	Upper limit
Maximum LST[a] (°C)	33	52	35	55
NDVI (arbitrary units)	1275	1550	1200	1600
Altitude (m)	500	1600	700	2700
Distance to nearest water body (km)	0	110	0	220

[a] LST: land surface temperature; NDVI: normalised difference vegetation index.

using the WHO recommended Kato-Katz technique[7] were included. Data were available from 1990–2009 in Oromia and from 1992–2009 in Western and Nyanza provinces.

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. 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).[8] To aid interpretation, semi-variance values were binned and averaged according to geographical separation distance. Subsequently, model semi-variograms were fitted through these binned values using a weighted least squares method. Where evidence of spatial autocorrelation exists, semi-variance typically rises with increasing 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 separation distance over which values are autocorrelated. 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.[9] 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.

The semi-variogram was then used to conditionally simulate 100 different, fully enumerated pseudo-datasets (termed realisations) at all 5695 schools in Western and Nyanza provinces in Kenya and all 5251 schools in Oromia Regional State in Ethiopia.[10] A population of 500 children was assumed at each school (a conservative estimate based on available data from Kenya Ministry of Education which suggests 420 children per primary 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 prevalences of 10% and 5%.

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 we used data from the National Oceanographic and Atmospheric Administration's Advanced Very High Resolution Radiometer, to derive estimates of maximum land surface temperature and normalised difference vegetation index[11] 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. Distance to permanent water bodies was derived in ArcMap 9.2 (ESRI, California, USA) from an electronic map obtained from the World Wildlife Fund.[12] All available data from the Global Atlas of Helminth Infection[5,6] for Kenya and Ethiopia 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% (Table 1). These thresholds were used to exclude schools from the sampling frame, reducing the number of schools: from 5695 to 4121 in Western and Nyanza and from 5251 to 4448 in Oromia (Figure 2).

2.4. Survey designs

Two sampling designs were considered: lot quality assurance sampling (LQAS)[13] and a variation of the lattice plus close pairs design (LpCP),[14] which provided a grid of points, with some additional pairs of points located close to each other, from which the prevalence values at unsurveyed schools were predicted using spatial interpolation methods. LQAS required all schools that lie within the ecological limits in each region/province to be sampled, whereas the LpCP 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.[8] Random sampling was not considered, as this has been shown to be less efficient for spatial prediction than a regular lattice design.[15]

The LQAS method allows the categorization of populations based on disease prevalence using small sample sizes for each sampling unit.[13,16] Previously, Brooker et al.[4]

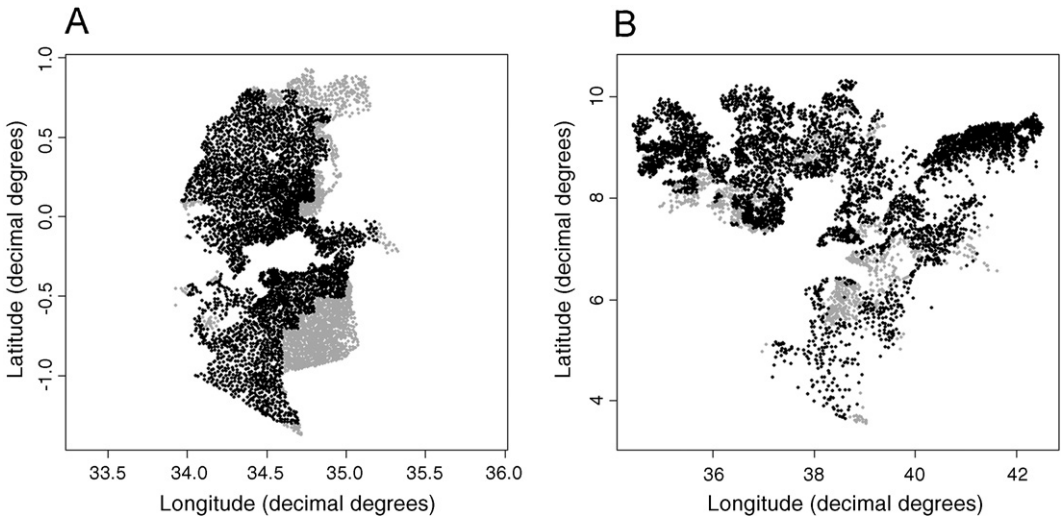

Figure 2. All public primary schools in A: Western and Nyanza provinces, Kenya; B: Oromia Regional State, Ethiopia. Schools shaded in gray indicate schools in areas of unlikely transmission.

used LQAS to categorize schools by *S. mansoni* prevalence in Uganda, whereby 15 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 between two and six samples 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%.[3] Using the simulated realisations of data for Western and Nyanza and for Oromia, we evaluated this sampling plan and an adapted plan using a lower stopping rule of only one positive.

The LpCP[14] approach surveys a subset of schools and uses the collected data and spatial interpolation methods to predict prevalence for all other unsurveyed schools. In order to make such predictions, the configuration of survey sites needed to be able to both estimate semi-variogram parameters and provide an efficient design from which to predict prevalence values at unsurveyed locations. The LpCP design balances both of these requirements and is based on a regular lattice with some additional close pairs of points (Figure 3).

For each lattice site (selected at the node of the grid), 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 50 additional survey sites in clusters of five, surrounding ten of the initial lattice sites (Figure 3). The inclusion of these additional sites allows for a more robust estimation of semi-variance

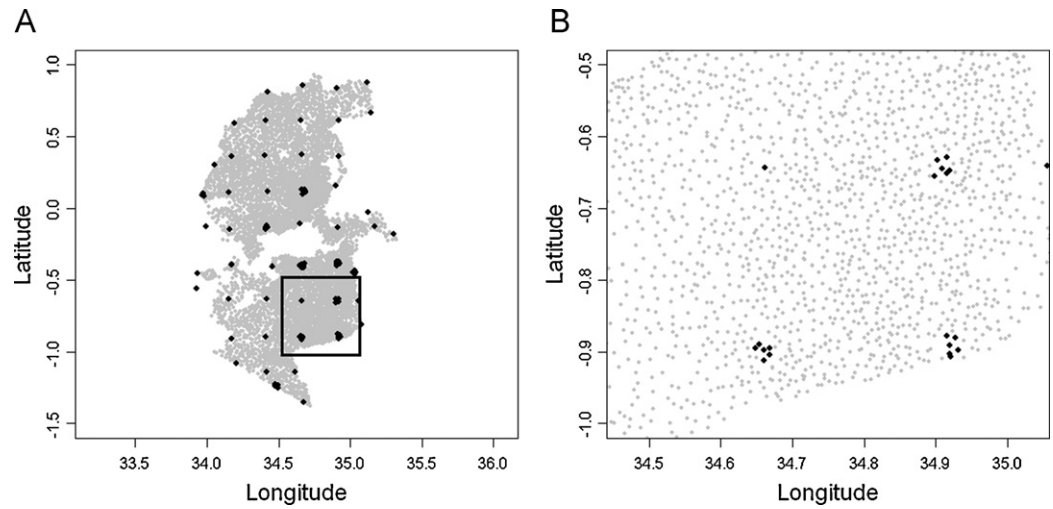

Figure 3. A: Illustrative example of the lattice plus close pairs design using a grid size of 27.5 km in Western and Nyanza provinces, Kenya. Dark points refer to survey schools and gray points to non-surveyed schools. B: Close-up of a region (black box in 3A) showing the locations of some of the clusters of closely located schools.

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, lattice sizes of smaller than 10 km resulted in sample sizes of greater than 3000 and therefore only lattices of 10–27.5 km were considered in this province. For each lattice size, 50 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 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 semivariogram 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.[8]

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) from 2008–2009. Relevant unit costs were identified according to an ingredients based approach.[17] 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 2). Two categories of costs were identified: imported equipment which was assumed to be similar in both settings and excluded costs of importation; and locally procured equipment, salary and transport costs, which were incurred locally and therefore differed between settings. Based on our field experience, we 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 75 km per day was assumed for both survey designs in Kenya, and 100 km 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.[18] 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$1 (May 2009) (www.oanda.com/convert/classic). To allow comparison, all costs were converted to 2009 US$ using the US$ Consumer Price Index (http://www.bls.gov/cpi/). All costs are expressed in 2009 US$. The effects of future inflation over the six years of the control programme were not included due to the difficulties in estimating future inflation rates in Kenya and Ethiopia.

Table 2
Itemized cost profile of *Schistosoma mansoni* school surveys in Kenya and Ethiopia in 2009 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 (US$)	
		Kenya	Ethiopia
Equipment[a]	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
Salaries (Fixed)[b]	Supervisor	28.57	42.55
	Technician	14.29	37.23
	Cleaner	7.14	5.32
Consumables (Fixed)[c]	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)[d]	Kato-Katz kits	0.30	0.30
	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
	Pencils	3.64	1.40
	Buckets	3.64	3.15
	Wash basins	3.64	3.15
Transport[e]	Transport (per day)	106.76	92.28

[a] Annuitized assuming a useful life of four years.
[b] Fixed cost per day.
[c] Fixed cost per school.
[d] Variable cost, dependent on number of children.
[e] Assumes an average distance of 75 km per day for Kenya and 100 km for 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, we used a recent estimate of combined delivery of praziquantel and albendazole to school-children,[19] and subtracted the reported unit costs of albendazole, 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 blanket treating all schools without carrying out surveys was also estimated. In addition, the total number of praziquantel treatments was estimated for each sampling strategy.

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;[20] (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).

2.7. Testing the performance of survey designs

Our primary performance metric is the proportion of schools requiring mass treatment (termed intervention schools) which are correctly classified. In addition, 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, we elected to use the total cost (survey plus treatment cost) per intervention school correctly classified 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.[21]

3. Results

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

3.1. Correct classification of schools

Against the derived gold standard data set, a LQAS plan using a stopping rule of two positives correctly classified 73.4% of intervention schools in Western and Nyanza and 74.3% in Oromia. A LQAS plan using a stopping rule of one positive, led to predictions that correctly classified 88.2% of intervention schools in Western and Nyanza and 89.5% in Oromia (Figure 5A). On the basis of these results, a sampling plan of using a stopping rule of one positive was used in the subsequent comparisons with the LpCP design.

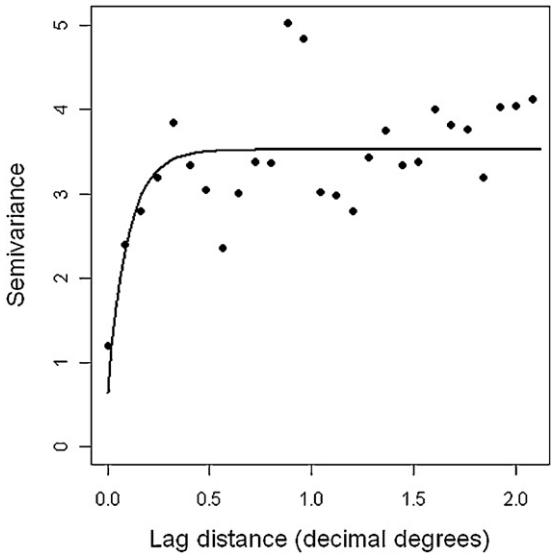

Figure 4. Semi-variogram of the prevalence of *Schistosoma mansoni* in 600 schools across Kenya and Ethiopia. Omnidirectional semi-variogram and best-fitted line of exponential spatial model for logistically transformed prevalence data is presented. Parameter values of the fitted spatial model were range = 0.31, sill = 3.52, nugget = 0.64. Directional semivariograms 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 110 km.

Figure 5A 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 27.5 km to 5 kms 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 smaller 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 in Oromia which, in turn, led to a higher proportion of intervention schools being correctly classified for a given grid size.

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, an LpCP design using a grid size of between 8 km in Kenya and 13.5 km in Ethiopia, correctly

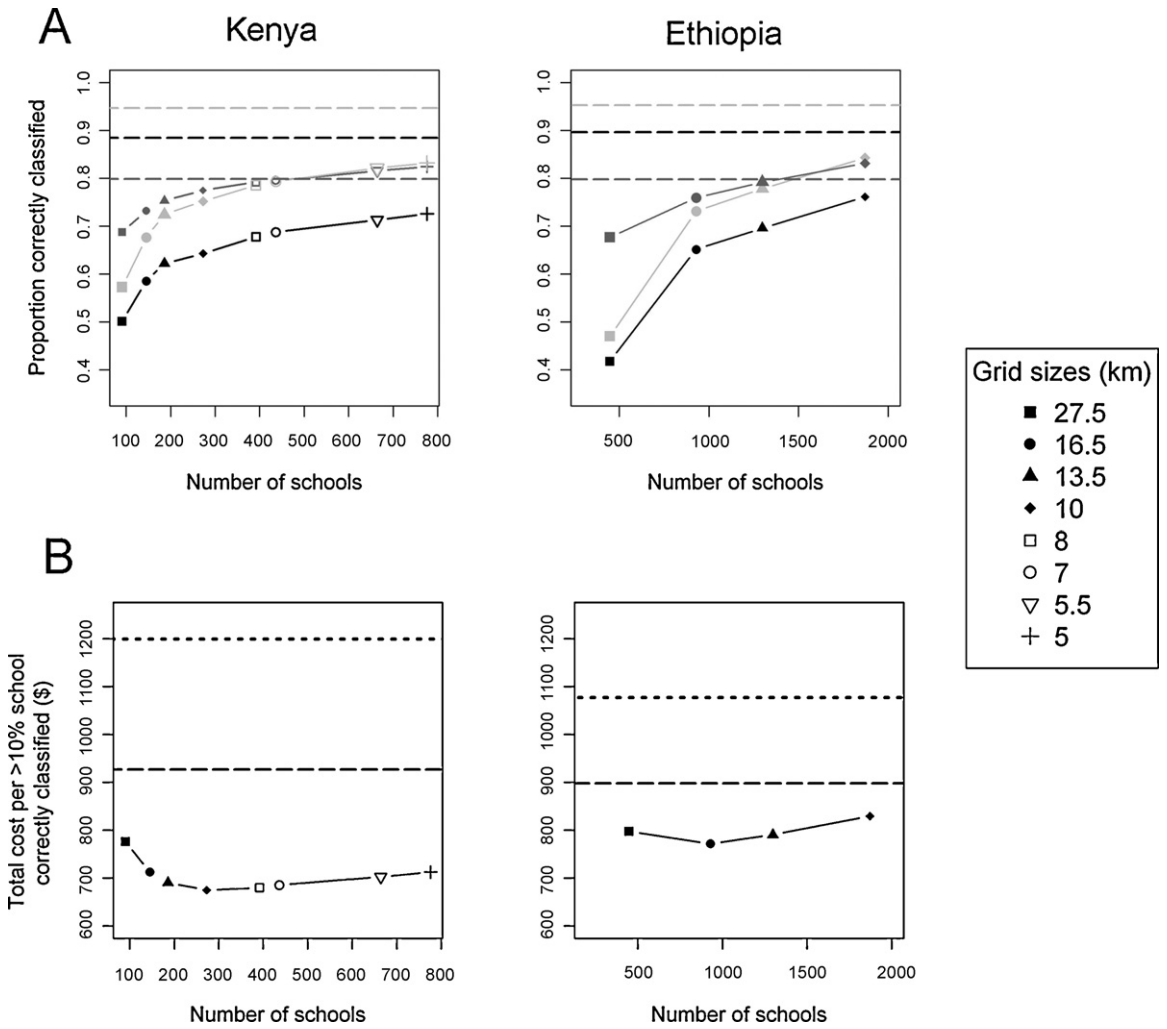

Figure 5. A: Proportion of intervention schools (where prevalence ≥10% and mass treatment is warranted) correctly classified using lot quality assurance sampling (LQAS) (black dashed line) and a lattice plus close pairs (LpCP) design (black solid line) for Western and Nyanza provinces, Kenya (left) and Oromia Regional State, Ethiopia (right). Light gray lines refer to the proportion of infected children within intervention schools correctly classified using LQAS (dashed) and LpCP (solid). Dark gray lines refer to the proportion of schools (prevalence < or ≥10%) correctly classified.
B: Cost-effectiveness of different survey designs using LQAS (dashed), a LpCP design (solid) and blanket treatment (dotted), in Western and Nyanza provinces, Kenya (left) and Oromia Regional State, Ethiopia (right). Black symbols denote the grid size, used in the LpCP design. Graphs assume six years of biennial treatment at a lower treatment cost of $0.295 per person. Note that the lines referring to blanket treatment are flat as no schools were surveyed using this approach. Similarly, lines referring to LQAS are flat as all schools were surveyed using this approach.

classified around the same proportion of schools as LQAS (Figure 5A).

3.2. Cost-effectiveness

Table 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 blanket treatment without applying an ecological mask. Whilst blanket treatment obviously requires no survey costs, the resource requirements in terms of praziquantel delivery are unfeasibly large in both study regions (Table 3). Use of either survey design resulted in lower overall cost than blanket 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 an LpCP design than LQAS. Likewise, treatment costs were generally lower when geostatistical designs were used, due to a higher proportion of intervention schools being wrongly classified as not requiring treatment.

Figure 5B 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 blanket treatment without surveys is less cost-effective than targeted treatment based on survey results. Another important result is that the LpCP design was generally more cost-effective than LQAS, irrespective of the number of schools that could be assessed

Table 3
Comparison of estimated total costs of surveys and consequent mass drug administration based on different survey approaches, assuming a drug delivery cost of $0.295 and six 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-efficient grid sizes are shown

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

LQAS: lot quality assurance sampling; LpCP: lattice plus close pairs.

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 5B - left). In Oromia (Figure 5B), 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.

3.3. *Sensitivity analysis*

The results of the sensitivity analyses are shown in Table 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 blanket 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 increasing the number of schools surveyed per day in LQAS in Ethiopia 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 distances 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.

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 rapid assessment method currently exists. The results suggest that implementing surveys to guide treatment delivery dramatically reduces both programme costs and the number of praziquantel treatments required. Current practice is for control programmes to deliver treatment to all schools within known *S. mansoni* endemic districts or sub-districts. The results further show that while LQAS correctly classifies a greater proportion of schools according to treatment requirement, the approach is more expensive than a geostatistical approach, which was shown to be more cost-effective in identifying schools with a high prevalence and warranting treatment.

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

Table 4
Sensitivity analysis of the cost-effectiveness (US$) of alternative sampling strategies in correctly classifying intervention schools in Western and Nyanza provinces, Kenya and Oromia Region, Ethiopia

Survey type	Baseline[a]	Higher drug cost	Higher fuel cost	LQAS (three schools/ day)	Economies of scale	Best case[c]	Worst case[c]
Western Kenya							
Blanket treatment	1199	1503	1235	1199	1019	1019	1548
LQAS	918	1072	982	830	836	746	1140
LpCP[b] (10 km)	675	831	700	675	583	583	861
LpCP (13.5 km)	690	855	714	690	593	593	884
Ethiopia							
Blanket treatment	1077	1351	1110	1077	916	916	1392
LQAS	907	1053	947	810	811	717	1097
LpCP (10 km)	772	924	806	772	682	682	964
LpCP (13.5 km)	790	935	829	790	705	705	978

[a] Baseline costs assume cheaper drug delivery cost of $0.295 and that two schools per day can be visited using lot quality assurance sampling (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%).

[b] Results from the most cost-effective efficient grid sized lattice plus close pairs (LpCP) design are shown for each country as well as 13.5 km grid.

[c] A best case scenario assumes lower drug cost, fixed fuel cost, 3 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.

programme 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 a consideration of practical and economic considerations as well as 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 we have highlighted in this study. Future work that links computer simulations 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 the shoreline of large water bodies, such as, for example, certain areas of Lake Victoria.[22–24] 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 we would encourage control programmes to make effective use of local expert knowledge to augment either LQAS or the geostatistical approach. Over large spatial scales, environmental data have been successfully integrated with geostatistical modeling to map the limits and broad patterns of schistosome transmission.[25–27] 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 a growing area of interest in ecological science[28] 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 Rapid Assessment of the Geographical Distribution of Bancroftian Filariasis (RAGFIL) method for lymphatic filariasis.[29] This approach recommends the selection of communities no more than 50 km apart to spatially interpolate a continuous estimate of prevalence over the study region.[30,31] The RAGFIL method has been used successfully to estimate the distribution of lymphatic filariasis in four countries in West Africa.[31] However, in addition to some concerns that small foci of infection may persist between interstices of a 50×50 km grid,[32] 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,[33,34] the cost-effectiveness of LQAS vs mass treatment[4] and surveys for soil-transmitted helminths.[10]

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 is restricted to four small foci: the lower Wabe Shebele valley, western Welega and lower and middle Awash valley.[35,36] In countries where *S. haematobium* is common, WHO recommends the use of blood in urine questionnaires, often implemented through the education system,[37] 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 a rapid *S. mansoni* survey to develop an overall national schistosomiasis control strategy.

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 spatial heterogeneity as observed in the present study.[26] Equally, semi-variograms estimated from survey data collected for Rwanda and Tanzania extracted from the Global Atlas of Helminth Infections indicates that *S. mansoni* appears to cluster at distances of 40–132 km (data not shown). Such consistency in clustering suggests that grid sizes of between 10–16.5 km would sufficiently capture the spatial heterogeneity of infection across sub-Saharan Africa. A further 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. However, work comparing *S. mansoni* infection in Mali suggested that in 2006, 12 years after the completion of a 10 year national drug campaign, the spatial distribution of infection was similar to that seen pre intervention.[38] 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. That said, an increasing number, perhaps even a majority, of ministries of education in Africa have georeferenced school databases as part of their Education Information Management System.

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 Province, a

larger grid size (16.5 km) is more cost-effective; whereas, where schools are more densely distributed, a finer grid (up to 10 km) seems more appropriate. As a compromise, we proposed a grid size of 13.5 km 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 requirement 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 increasing availability of open-access spatial tools, such as open source GIS software and R statistical package,[21] this will become an increasingly viable option. Importantly, adding the cost of any external technical assistance (approximately US$ 10 000) would not change the overall conclusions of the study.

In summary, using a computerized simulation approach, we have shown that targeting praziquantel at school/community level is more cost-effective than blanket treatment for the control of *S. mansoni*. We further show that while LQAS correctly classifies a greater proportion of schools according to treatment requirement, a geostatistical approach is more cost-effective in identifying high prevalence schools. 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.

Authors' contributions: HJWS and SB conceived the study, HJWS analysed and interpreted data, PWG contributed to data analysis and interpretation, SB, JHK, RAA and NBK collected and interpreted data, HJWS drafted the manuscript and SB assisted with revisions. All authors read and approved the final manuscript.

Acknowledgements: We would like to thank all those involved in the collection of the data used in this study and Rachel Pullan for comments.

Funding: HJWS is supported by a London School of Hygiene and Tropical Medicine Graduate Teaching Assistantship, PWG is supported under a Wellcome Trust Senior Research Fellowship held by Dr Simon Hay at the University of Oxford (#079091), and SB is supported by a Research Career Development Fellowship from the Wellcome Trust (#081673). This work forms part of the output of the Global Atlas of Helminth Infection which is principally funded by the Partnership for Child Development at Imperial College UK, Deworm the World, and the Wellcome Trust.

Conflicts of interest: None declared.

Ethical Approval: Not required.

References

1. Lengeler C, Utzinger J, Tanner M. Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa. *Bull World Health Organ* 2002;**80**:235–42.
2. Brooker S, Kabatereine NB, Gyapong JO, Stothard JR, Utzinger J. Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology* 2009;**136**:1707–18.
3. WHO. *Preventive chemotherapy in human helminthiasis. Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*. Geneva: World Health Organization; 2006.
4. Brooker S, Kabatereine NB, Myatt M, Stothard RJ, Fenwick A. Rapid assessment of *Schistosoma mansoni*: the validity, applicability and cost-effectiveness of the Lot Quality Assurance Sampling method in Uganda. *Trop Med Int Health* 2005;**10**:647–58.
5. Brooker S, Kabatereine NB, Smith JL, Mupfasoni D, Mwanje MT, Ndayishimiye O, et al. An updated atlas of human helminth infections: the example of East Africa. *Int J Health Geog* 2009;**8**:42.
6. Brooker S, Rowlands M, Haller L, Savioli L, Bundy DAP. Towards an atlas of human helminth infection in sub-Saharan Africa: The use of Geographical Information Systems (GIS). *Parasitol Today* 2000;**16**:303–7.
7. WHO. *Basic laboratory methods in medical parasitology*. Geneva: World Health Organization; 1991.
8. Goovaerts P. Geostatistics for natural resources evaluation. New York: Oxford University Press; 1997.
9. Cressie N. Geostatistical methods for mapping environmental exposures. In: Elliott P, Wakefield JC, Best NG, Briggs DJ, editors. *Spatial epidemiology, methods and applications*.. Oxford: Oxford University Press; 2000. p. 185–204.
10. Sturrock HJW, Gething PW, Clements AC, Brooker S. Optimal survey designs for targeting chemotherapy against soil-transmitted helminths: effect of spatial heterogeneity and cost-efficiency of sampling. *Am J Trop Med Hyg* 2010;**82**:1079–87.
11. Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. Global environmental data for mapping infectious disease distribution. *Adv Parasitol* 2006;**62**:37–77.
12. Olson DM, Dinerstein E. The Global 200: Priority ecoregions for global conservation. *Ann Mo Bot Gard* 2009;**89**:125–6.
13. Lemeshow S, Taber S. Lot quality assurance sampling: single- and double-sampling plans. *World Health Stat Q* 1991;**44**:115–32.
14. Diggle PJ, Lophaven S. Bayesian geostatistical design. *Scand J Stat* 2006;**33**:53–64.
15. Diggle PJ, Ribeiro Jnr PJ. Model-based geostatistics. New York: Springer; 2007.
16. Robertson SE, Valadez JJ. Global review of health care surveys using lot quality assurance sampling (LQAS), 1984–2004. *Soc Sci Med* 2006;**63**:1648–60.
17. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.
18. Walker D, Kumaranayake L. Allowing for differential timing in cost analyses: discounting and annualization. *Health Policy Plan* 2002;**17**:112–8.
19. Gabrielli AF, Toure S, Sellin B, Sellin E, Ky C, Ouedraogo H, et al. 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 Trop* 2006;**99**:234–42.
20. Brooker S, Kabatereine NB, Fleming F, Devlin N. Cost and cost-effectiveness of nationwide school-based helminth control in Uganda: intra-country variation and effects of scaling-up. *Health Policy Plan* 2008;**23**:24–35.

21. R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2008.
22. Lwambo NJS, Siza JE, Brooker S, Bundy DAP, Guyatt H. Patterns of concurrent infection with hookworm and schistosomiasis in school children in Tanzania. *Trans R Soc Trop Med Hyg* 1999;**93**: 497–502.
23. Brooker S, Miguel EA, Waswa P, Namunyu R, Moulin S, Guyatt H, et al. The potential of rapid screening methods for *Schistosoma mansoni* in western Kenya. *Ann Trop Med Parasitol* 2001;**95**:343–51.
24. Handzel T, Karanja DM, Addiss DG, Hightower AW, Rosen DH, Colley DG, et al. Geographic distribution of schistosomiasis and soil-transmitted helminths in Western Kenya: implications for anthelminthic mass treatment. *Am J Trop Med Hyg* 2003;**69**:318–23.
25. Simoonga C, Utzinger J, Brooker S, Vounatsou P, Appleton CC, Stensgaard AS, et al. Remote sensing, geographical information system and spatial analysis for schistosomiasis epidemiology and ecology in Africa. *Parasitology* 2009;**136**:1683–93.
26. Brooker S. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Trans R Soc Trop Med Hyg* 2007;**101**:1–8.
27. Magalhães RJS, Clements ACA, Patil AP, Gething PW, Brooker S. The applications of model-based geostatistics in helminth epidemiology and control. *Adv Parasitol* 2011;**74**:267–96.
28. Hirzel A, Guisan A. Which is the optimal sampling strategy for habitat suitability modelling. *Ecol Model* 2002;**157**:331–41.
29. WHO. *Operational guidelines for rapid mapping of Bancroftian filariasis in Africa*. Geneva: World Health Organization; 2000.
30. Gyapong JO, Remme JH. The use of grid sampling methodology for rapid assessment of the distribution of bancroftian filariasis. *Trans R Soc Trop Med Hyg* 2001;**95**:681–6.
31. Gyapong JO, Kyelem D, Kleinschmidt I, Agbo K, Ahouandogbo F, Gaba J, et al. The use of spatial analysis in mapping the distribution of bancroftian filariasis in four West African countries. *Ann Trop Med Parasitol* 2002;**96**:695–705.
32. Srividya A, Michael E, Palaniyandi M, Pani SP, Das PK. A geostatistical analysis of the geographic distribution of lymphatic filariasis prevalence in southern India. *Am J Trop Med Hyg* 2002;**67**:480–9.
33. Connelly LB. Balancing the number and size of sites: an economic approach to the optimal design of cluster samples. *Control Clin Trials* 2003;**24**:544–59.
34. Williams B, Gopi PG, Borgdorff MW, Yamada N, Dye C. The design effect and cluster samples: optimising tuberculosis prevalence surveys. *Int J Tuberc Lung Dis* 2008;**12**:1110–5.
35. Kloos H, Lo CT, Birrie H, Ayele T, Tedla S, Tsegay F. Schistosomiasis in Ethiopia. *Soc Sci Med* 1988;**26**:803–27.
36. WHO. *Atlas of the Global Distribution of Schistosomiasis*. Geneva: WHO; 1987.
37. Clements AC, Brooker S, Nyandindi U, Fenwick A, Blair L. Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *Int J Parasitol* 2008;**38**:401–15.
38. Clements AC, Bosque-Oliva E, Sacko M, Landoure A, Dembele R, Traore M, et al. A comparative study of the spatial distribution of schistosomiasis in Mali in 1984–1989 and 2004–2006. *PLoS Neg Trop Dis* 2009;**3**:e431.