

## Optimal Survey Designs for Targeting Chemotherapy Against Soil-Transmitted Helminths: Effect of Spatial Heterogeneity and Cost-Efficiency of Sampling

Hugh J. W. Sturrock,\* Peter W. Gething, Archie C. A. Clements, and Simon Brooker

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom; Spatial Ecology and Epidemiology Group, Department of Zoology, University of Oxford, Oxford, United Kingdom; School of Population Health, University of Queensland, Herston, Queensland, Australia; Australian Centre for International and Tropical Health, Queensland Institute of Medical Research, Herston, Queensland, Australia; Malaria Public Health and Epidemiology Group, Kenya Medical Research Institute–Wellcome Trust Research Programme, Nairobi, Kenya

**Abstract.** Implementation of helminth control programs requires information on the distribution and prevalence of infection to target mass treatment to areas of greatest need. In the absence of data, the question of how many schools/communities should be surveyed depends on the spatial heterogeneity of infection and the cost efficiency of surveys. We used geostatistical techniques to quantify the spatial heterogeneity of soil-transmitted helminths in multiple settings in eastern Africa, and using the example of Kenya, conducted conditional simulation to explore the implications of alternative sampling strategies in identifying districts requiring mass treatment. Cost analysis is included in the simulations using data from actual field surveys and control programs. The analysis suggests that sampling four or five schools in each district provides a cost-efficient strategy in identifying districts requiring mass treatment, and that efficiency of sampling was relatively insensitive to the number of children sampled per school.

### INTRODUCTION

In the past two decades, there has been an expansion in the implementation of programs to control a range of helminth species using population-based chemotherapy.<sup>1</sup> Among the most prevalent helminth species in humans are soil-transmitted helminths (STHs), including *Ascaris lumbricoides*, *Trichuris trichiura*, and the hookworms *Ancylostoma duodenale* and *Necator americanus*. Mass drug administration (where all persons are treated regardless of infection status) using benzimidazole anthelmintics is recommended for communities where STH infection prevalence exceeds 20%.<sup>2</sup> Identifying these communities requires reliable information on the prevalence and geographic distribution of infection.

To address this operational requirement there has been increased interest in investigating scientifically robust yet practical approaches to mapping.<sup>3</sup> It has been demonstrated that climate-based risk models can reliably define the large-scale limits of transmission,<sup>4</sup> providing an initial stage in the geographic targeting of treatment. Within these broad limits, epidemiologic surveys are still required to identify localities requiring mass treatment. However, large-scale STH surveys increase the cost of any program and may prove difficult for many developing countries to implement because of a lack of technical and financial resources.<sup>5,6</sup> Similar operational challenges exist for control programs targeting other tropical diseases including filariasis, onchocerciasis, schistosomiasis and trachoma. In response, efforts have been made to reduce the cost and complexity of surveys for targeting treatment by developing survey tools and methods that enable rapid assessment of health problems.<sup>7–9</sup> Examples include blood in urine questionnaires for urinary schistosomiasis,<sup>10</sup> lot quality assurance sampling (LQAS) for intestinal schistosomiasis,<sup>11,12</sup> trachoma,<sup>13</sup> and malaria,<sup>14</sup> and rapid assessment methodologies for filariasis,<sup>15</sup> onchocerciasis<sup>16</sup> and loiasis.<sup>17</sup>

To date, no study has investigated the utility of rapid assessment procedures to support the control of STHs. 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 drug procurement and delivery required for comprehensive treatment strategies. Second, STHs have been assumed to be geographically homogeneous, with similar infection levels occurring over large distances.<sup>18</sup> Spatial heterogeneity has important consequences for surveys and control because it determines the resolution at which surveys and interventions need to be carried out. Diseases that are widespread and evenly distributed 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 STHs at a scale of operational relevance.

Defining an optimal sampling scheme for targeting STH control requires an understanding of the following issues: 1) the degree of spatial heterogeneity of STH infection; 2) the financial and human cost of conducting epidemiologic surveys for STHs; 3) the geographic framework within which public health decision-making is organized through community, sub-district and district levels; and 4) the financial and public health consequences of inappropriate control decisions on the need for mass treatment. The aim of the present study was 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 to classify intervention units according to treatment strategy.

### METHODS

**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, which enabled comparison of results over

\* Address correspondence to Hugh J. W. Sturrock, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, 51 Bedford Square, London, WC1B 3DP, United Kingdom. E-mail: hugh.sturrock@lshtm.ac.uk

a range of transmission and ecologic settings. Using the example of Kenya, these spatial characteristics were used to parameterize simulation analyses that explored the implications of spatial survey design for enumerating district-level (second 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 because of the availability of a national, georeferenced schools database and detailed, standardized survey data from Kenya and from bordering areas supporting empirical estimates of spatial heterogeneity.

**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 methods (i.e., examination of school age children by using Kato-Katz examination of stool samples)<sup>19</sup> for coastal Kenya (Brooker S and others, unpublished data), Uganda,<sup>20</sup> northwestern Tanzania,<sup>21</sup> and Zambia (Mwanza J and others, unpublished data) (Figure 1). These surveys represent some of the most geographically comprehensive survey data for STHs in sub-Saharan Africa.

**Geostatistical analyses.** Spatial heterogeneity of infection prevalence was investigated by using semi-variogram analysis. A semi-variogram characterizes 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).<sup>22,23</sup> 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.<sup>24</sup> 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 by inspection of raw prevalence variograms, data were detrended by 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.<sup>4</sup> The resultant normally distributed Pearson residuals were used to estimate the semi-variogram. In the remaining countries where no evidence of large-scale trends was detected, because of 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 by using weighted least squares fits of exponential, spherical, and gaussian models and examined by visual inspection.<sup>22</sup> Analyses were carried out by using the GeoR package in R 2.7.1.<sup>25</sup> Semivariograms were generated for each country separately, except for Tanzania and Uganda, which represented contiguous areas and were therefore analyzed initially as a single dataset. Because of 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.

**Conditional simulation and cost analysis: case study of Kenya.** The exploration of spatial heterogeneity enabled the generation of a pseudo-data set that had the same spatial and variance characteristics as those expected in the field. Simulating a completely enumerated dataset enabled different sampling strategies to be evaluated against a realistic gold standard. To achieve such a gold standard, we used conditional simulation that uses parameters arising from the variogram analysis to generate a range of different scenarios (or realizations) 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

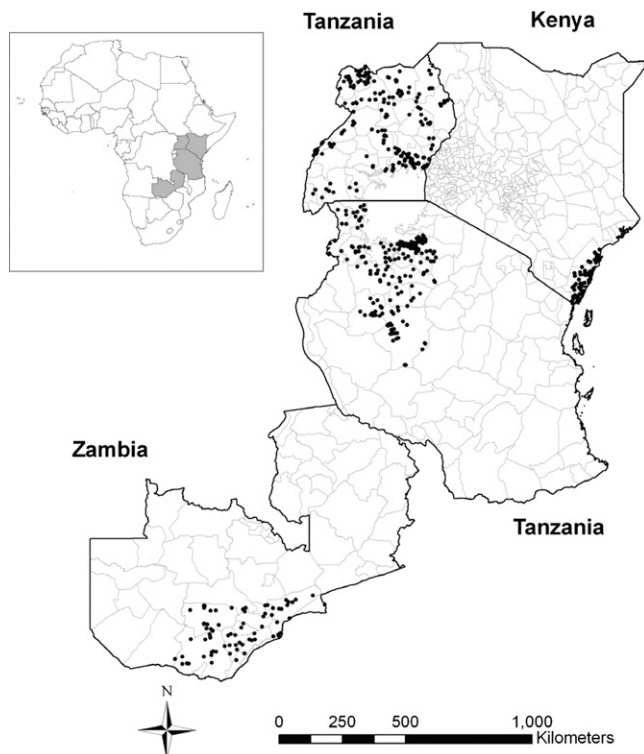


FIGURE 1. Locations of the countries included in the study (inset map) and locations of schools from the datasets included in the study (main map).

and a mean and variance are predicted by 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, one can select a random value 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 realization. As the final set of prevalence values in a realization is dependent on the order of selection of prediction locations and the values assigned at each location, different realizations are unique in terms of locations of clusters and overall prevalence.<sup>26</sup> Data were simulated for all government mixed primary schools in Coast, Nyanza and Western provinces, the most populous provinces of Kenya and where STHs are most prevalent,<sup>27</sup> using the Kenya data and variograms. Information on the schools and their location was obtained from the Ministry of Education school database (Figure 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 12 districts in Nyanza Province. To enable sampling designs to be tested on contrasting scenarios, data were also conditionally simulated for schools in Western and Nyanza provinces by using data and variograms from neighboring Uganda.

For each STH species and in each region, 1,000 realizations 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 formula  $p = H + A + T - (HA) - (HT) - (AT) + (HAT)$ ,<sup>28</sup> 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 1, the expected prevalence of co-infection with different species combinations was calculated assuming complete independence<sup>28</sup> and compared with 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 were 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 that suggest 420 children per primary school), we considered the following sampling strategies: 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 realization, 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 World Health Organization endemicity classes (low = < 20%, medium =  $\geq 20\%$  to < 50%, high =  $\geq 50\%$ ).<sup>29</sup> These classifications were compared with 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 was 7,861 km<sup>2</sup> (range = 232–38,701 kms) in Coastal Province, 959 km<sup>2</sup> (range = 581–1,994 km<sup>2</sup>) in Nyanza Province, and 937 km (range = 556–2,058 km<sup>2</sup>) in Western Province. All the above simulations were carried out by using bespoke scripts written in R 2.7.1.

The cost of each sampling strategy was estimated by using itemized costs collected for the Kenya survey. In accordance with 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 shown in Table 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, we assumed that one supervisor, one technician, and one cleaner were

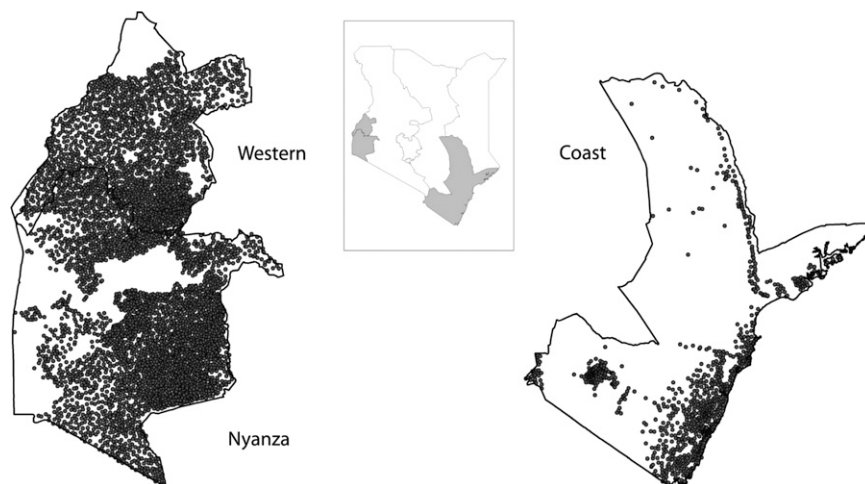


FIGURE 2. Map of Kenya showing province boundaries (**inset map**) and locations of 6,899 mixed public primary schools in Western, Nyanza, and Coast provinces (**main map**). The size and position of Western and Nyanza provinces are not in relation to Coast Province.

TABLE 1  
Summary of data on soil-transmitted helminth infection used in the geostatistical analysis of spatial heterogeneity

Country	No. of schools	Median prevalence (%) (range by school)			Mean no. examined	Mean age examined (years)	Reference
		Hookworm	<i>Ascaris lumbricoides</i>	<i>Trichuris trichiura</i>			
Kenya	64	16.7 (0–69.2)	0 (0–23.4)	4.7 (0–54.4)	104.3	12.2	Brooker, unpublished data
Tanzania	197	43.3 (1.6–93.3)	0 (0–40)	0 (0–43.3)	75.9	14.03	Clements and others <sup>21</sup>
Uganda	197	52.4 (0–90)	0 (0–65.8)	0 (0–40)	68.9	10.21	Kabatereine and others <sup>20</sup>
Zambia	79	11.1 (0–86.7)	0 (0–333)	0 (0–21.3)	59.8	13.06	Mwanza and others, unpublished data

required per day, irrespective of the number of children sampled (category 2). Where 31–60 children were sampled per school, an extra technician was included and if ≥ 60 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 km/day was assumed and a 10% contingency allowance was also included. On the basis of recent field experience in Kenya, we also assumed that one school could be surveyed per day. Capital costs (category 1) were annuitized over the useful life of each item by using a discount rate of 3%, which is consistent with the recommendations of the World Bank.<sup>30</sup> 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.<sup>31</sup> Vehicle running costs only included maintenance and insurance. Costs were estimated in

local currency and their current values were converted into equivalent US\$ by September 1, 2008 exchange rates of 70.25 Kenyan Shillings to 1US\$ and 0.55 British Pounds to 1US\$ ([www.oanda.com/convert/classic](http://www.oanda.com/convert/classic)).

In addition, the cost of treatment and delivery was calculated by using two recent estimates of \$0.15 and \$0.39 per delivery round per child.<sup>32,33</sup> 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 mass drug administration that would be carried out based on results of that survey. By including 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 implications of each sampling strategy, the cost per district correctly classified was calculated by dividing the average total cost of each sampling strategy across the 1,000 realizations by the average number of districts correctly classified per realization. For the purpose of this study, we refer to the cost per district correctly classified as the cost efficiency of a sampling strategy.

TABLE 2  
Itemized costs of conducting school surveys of soil-transmitted helminth infections in Kenya\*

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)†	Supervisor	28.57	4
		Technician	14.29	4
Cleaner		7.14	4	
Salaries (variable)‡	Technician	14.29	0–8	
Consumables (fixed)§	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)¶	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#		Fuel and maintenance (per km)	1.43	300

\*The number of units required for one district using four schools are shown with ranges according to 10–60 children per school.

† Fixed cost per day.

‡ Variable cost dependent on number of children.

§ Fixed cost per day.

¶ Variable cost dependent on number of children.

# Assumes an average distance of 75 km/day.

RESULTS

Prevalence data were available for 537 schools and 39,924 children. The median prevalence of hookworm ranged from 11.1% to 52.4% between countries, and the median prevalence for *A. lumbricoides* across all countries was 0%, and the prevalence for *T. trichiura* ranged from 0% to 4.7% (Table 1).

Species-specific semi-variograms for each study region and distinct differences in the degree of spatial heterogeneity for hookworm compared with *A. lumbricoides* and *T. trichiura* are shown in Figure 3. Specifically, the semi-variograms for hookworm indicate spatial autocorrelation with fitted range parameters between approximately 95 km and 166 km. In contrast, the semi-variograms for *A. lumbricoides* showed either no spatial autocorrelation or spatial autocorrelation with shorter fitted range parameters between 36 km and 92 km. Similarly, the semi-variograms for *T. trichiura* only indicated spatial autocorrelation in three of the datasets, with ranges between 44 km and 46 km, and other datasets showed little evidence of spatial autocorrelation. These results indicate that in eastern Africa there is a consistency in the scale over which species-specific spatial autocorrelation occurs, and that spatial autocorrelation in hookworm prevalence occurs over much larger distances than that for *A. lumbricoides* and *T. trichiura*.

In Kenya, simulations based on data and variograms for Coast Province yielded cumulative STH prevalence estimates ranging from 28% to 42% in Coast Province, and from 12% to 65% in Western and Nyanza provinces. The wider range of prevalence values simulated for Western and Nyanza



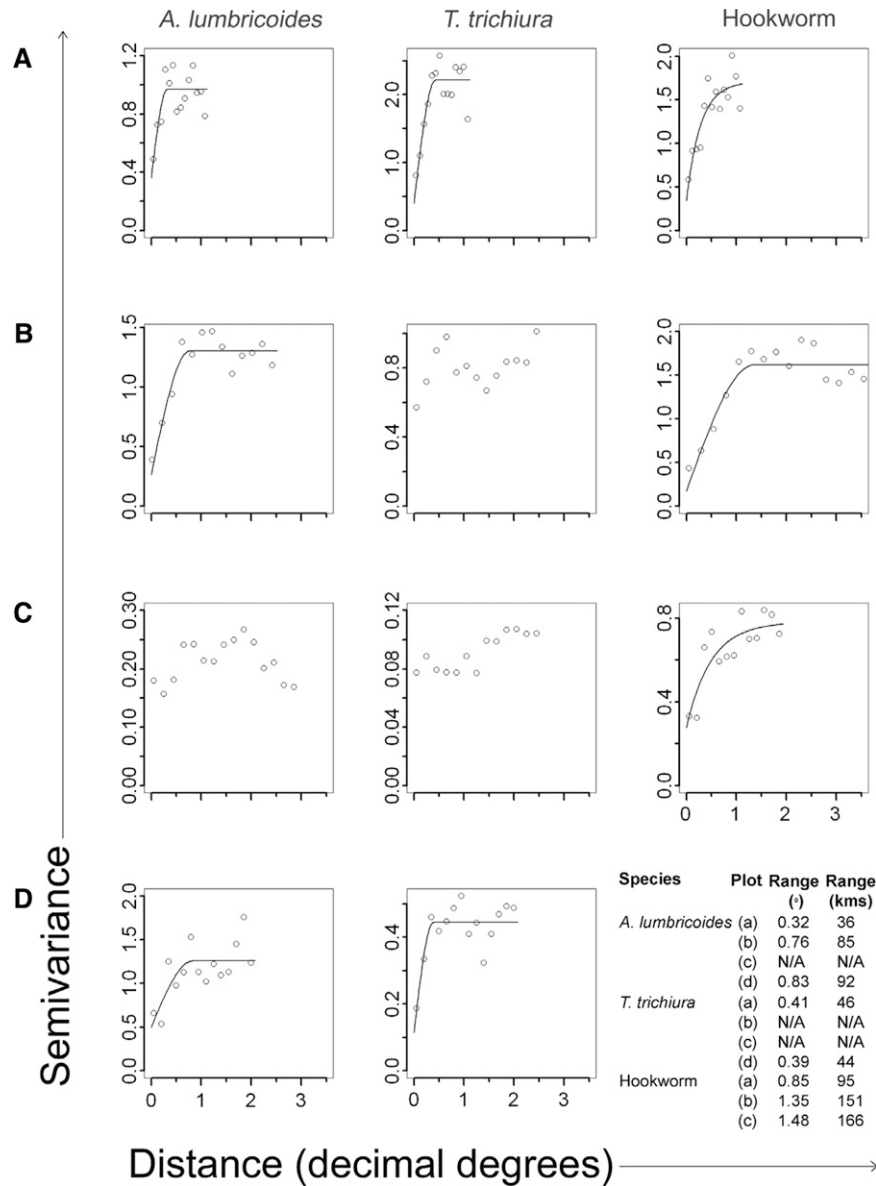


FIGURE 3. Semi-variograms and best-fitted lines of spatial models for soil-transmitted helminths *Ascaris lumbricoides* and *Trichuris trichiura* in **A**, Kenya; **B**, Uganda and northern Tanzania > 2° south; **C**, Uganda and northern Tanzania < 2° south; and **D**, Zambia; and hookworm in **A**, Kenya; **B**, Uganda and Tanzania; and **C**, Zambia. Vertical axis indicates semi-variance and horizontal axis indicates distance in decimal degrees. Range in kilometers was calculated assuming 1 decimal degree = 112 km at the equator.

provinces is the result of the higher degree of uncertainty caused by the lack of survey points in this area. Using data and variograms for Uganda, we found that simulations yielded cumulative STH prevalences ranging from 27% to 90% for Western Province and from 20% to 90% for Nyanza Province.

The trade-off between the number of schools surveyed and the proportions of districts correctly classified in each province, averaged over all sample sizes at each school, is shown in Figure 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. Thus, sampling more than four schools yields little extra performance. The range in performance by using 10 and 60 children per school and 4 schools per district is shown in Table 3.

Altering the number of children sampled in each school made little difference in overall accuracy.

Sensitivity analysis found that the prevalence of co-infection in the 537 schools (Table 1) was slightly higher than would be expected by chance, with co-infection with hookworm and *Ascaris* being 1.1 times higher than expected by chance, co-infection with hookworm and *Trichuris* being 1.13 times higher, co-infection with *Ascaris* and *Trichuris* being 1.3 times higher, and co-infection with all three species being 1.7 times higher. Use of these probabilities led to lower estimates of cumulative STH prevalence at each school including the simulations. However, these different estimates made little difference in performance of sampling strategies (Table 3).

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

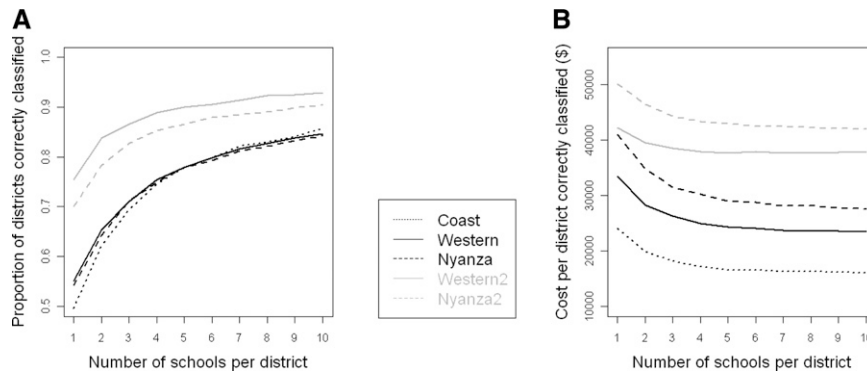


FIGURE 4. **A**, Relationship between the number of schools surveyed and the ability to correctly classify districts according to treatment strategy in Coast, Nyanza, and Western provinces, based on conditionally simulated data from Kenya data and variograms (Coast, Western 1 and Nyanza 1) and Uganda data and variograms (Western 2 and Nyanza 2). Results are averaged over the different numbers of children sampled per school because there was little effect of sample size. **B**, Cost implications of different sampling strategies, averaged across different number of children per school, using a treatment cost of \$0.15 per person and considering one year of treatment.

cost to survey one school ranged from \$192 when 1 school and 60 children per school were surveyed per district to \$302 when 10 schools and 10 children per school were surveyed per district. The survey cost per school ranged from \$191 to \$295 in Western Province and from \$189 to \$277 in Nyanza Province. The relationship between the number of schools surveyed (averaged over the different numbers of children sampled at each school) and the total (survey and treatment) costs per district correctly classified assuming a treatment cost of \$0.15 per person per round and one year of treatment is shown in Figure 4B. 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 leads to large cost savings, whereas surveying  $\geq 4$ –5 schools per district results in diminishing returns in terms of cost savings. Varying treatment cost and delivery time period yielded similar conclusions. As found with performance, increasing the number of children surveyed per school made little difference to cost-efficiency: 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 the number of children sampled per school was increased from 10 to 60.

Survey costs and cost of misclassification (cost of unnecessary treatment) for each sampling strategy using Coast Province and treatment costs of \$0.15 over 1 year as an example are shown in Figure 5. Although survey costs increase linearly, the cost of misclassification decreases non-linearly,

which reflected the non-linear increase in accuracy associated with increasing sample sizes.

DISCUSSION

Central to the implementation of cost-effective helminth control is the need to target mass treatment to areas of greatest need. To reduce program costs, surveys to guide such targeting should be reliable but also rapid and low cost.<sup>9</sup> This study represents a first attempt to examine the spatial heterogeneities of STH infection and its implications when optimizing 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, the scale of spatial autocorrelation is generally similar across different transmission settings. Using the case study of Kenya, we showed by simulation studies 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 diminishing returns in performance to correctly classify districts and the cost per district correctly classified.

The results of geostatistical analyses corroborate with those of a previous study in Uganda,<sup>18</sup> 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

TABLE 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\*

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	60 children	10 children	60 children	10 children	60 children	10 children	60 children
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

\* 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 1.

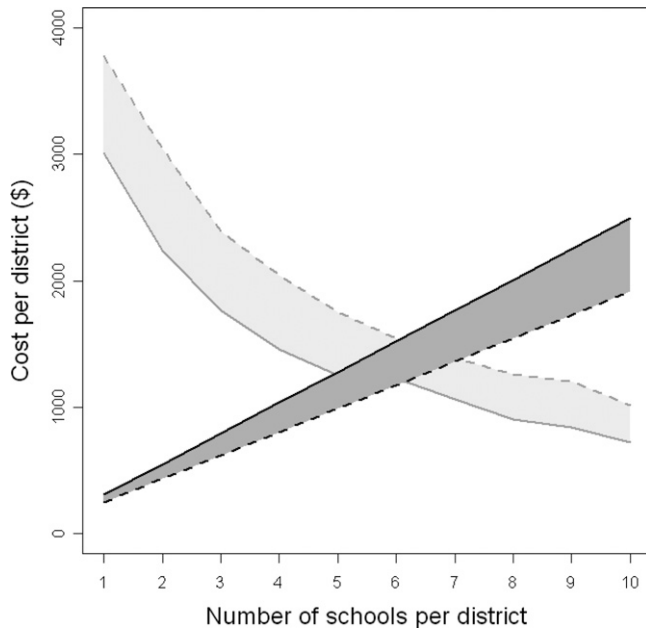


FIGURE 5. Cost per district of the different sampling strategies (dark gray) and the cost of unnecessary treatment in those districts incorrectly classified in a higher endemicity class and requiring mass treatment, when mass treatment was not required (light gray) in Coast Province, Kenya, assuming treatment costs of \$0.15 over one year. The solid lines represent 60 children per school and the dashed lines represent 10 children per school.

autocorrelation observed for hookworm suggest that spatially structured variables other than those included in the regression model affect hookworm transmission; possibly soil type.<sup>34,35</sup> 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. Due to the more widespread distribution of hookworm,<sup>4</sup> STH infections collectively are less focal than either schistosomiasis or filariasis, which show autocorrelation up to distances < 50 km.<sup>36–38</sup> The requirement of an intermediate host for schistosomiasis and a vector for lymphatic filariasis 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, which enable transmission where environmental conditions suit free-living parasite stages.<sup>39</sup>

Such inherent differences in the spatial heterogeneity have important implications for the design of integrated surveys that simultaneously survey STHs, schistosomiasis, and filariasis. The more widespread distribution of STHs implies that STH surveys can readily be integrated into survey for schistosomiasis and filariasis because the spatial sampling method developed for these two diseases will sufficiently capture the spatial heterogeneities of STH infection. Current recommendations for lymphatic filariasis suggest a maximum of two sites per district should be surveyed to assess whether prevalence is > 1%, although there is debate as to whether this sampling strategy and the 50 km × 50 km grid-based rapid assessment,<sup>40</sup> have a sufficiently fine-scale spatial resolution to capture foci of infection.<sup>37,38</sup>

Studies investigating the costs of surveys in the developing world are surprisingly sparse. A number of studies have evaluated the cost of screening persons for helminth infections<sup>41,42</sup>

and screening versus presumptive anthelmintic treatment.<sup>11,43</sup> For tuberculosis, Williams and others investigated the trade-off between sampling effort and survey cost for clustered survey designs in Cambodia.<sup>44</sup> 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 increasing as more clusters are sampled.<sup>44</sup> The present study is, to our knowledge, the first to evaluate the cost implications of different sampling strategies to guide treatment strategies against STHs and 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-efficient sampling method to guide STH control in eastern Africa. Although analysis suggests that surveying up to 10 schools per district has the greatest cost-efficiency, this benefit was marginal and surveying four to five schools provides a balance between operational ease and cost-efficiency. In addition, because increasing the number of children surveyed at each of these schools from 10 to 60 makes little difference to overall cost-efficiency of sampling, relatively small numbers of children per school provide a cost-efficient strategy. However, if the aim of the survey was to estimate prevalence, the sample size would influence precision.<sup>45</sup> Evaluation of lot quality assurance sampling (LQAS) for the rapid assessment of *Schistosoma mansoni* in Uganda showed that sampling 15 children per school provided a cost-effective strategy to guide the targeting of praziquantel treatment of schistosomiasis.<sup>11</sup> Further evaluation of LQAS for STH infection is warranted. Finally, the relatively large distances over which spatial autocorrelation for hookworm occurs implies that sampling strategy developed for the more spatial focal schistosomiasis and lymphatic filariasis will capture the spatial heterogeneities in hookworm, the more widespread STH species in much of Africa.<sup>4</sup>

Although our study provided a thorough examination of different sampling schemes, it is important to highlight some of the limitations of this study. First, for conditional simulation, due to the lack of empirical estimates, we assumed that the spatial processes that occur in Western and Nyanza provinces are equal to those found in either coastal Kenya or Uganda. Second, we have also assumed that the probability of coinfection, both geographically and at the individual level, is independent.<sup>28</sup> Third, it is likely that the spatial heterogeneity of STH differ in equatorial western Africa, where hookworm is often less prevalent than either *A. lumbricoides* and *T. trichiura*.<sup>4</sup> Further geostatistical analyses and an exploration of sampling designs in this region are required to better understand these issues. Fourth, most of the data used for the semi-variogram analysis were collected before implementation of large-scale treatment programs. It is nevertheless possible that small-scale treatment in specific, unknown locations may have altered the spatial heterogeneity of infection. However, a comparative analysis of the spatial heterogeneity of schistosomiasis in Mali between 1984 and 1989 and 2004 and 2006 found consistent patterns of spatial autocorrelation<sup>46</sup>; no comparable analysis has been undertaken to date for STH species.

In conclusion, we have provided an initial quantification of the spatial heterogeneity of STH over a number of settings in eastern Africa and show that hookworm consistently exhibits

spatial autocorrelation over larger distances than either *A. lumbricoides* or *T. trichiura*. The implication of this result is that in areas where hookworm is more widespread, such as in eastern Africa,<sup>27</sup> decisions about the need to provide mass treatment can confidently be made at the district level. We further show that sampling small numbers of children in four to five schools in each district provides robust, quick, and cost-efficient sampling strategies to identify districts requiring mass treatment in an east African setting. Further work is required to investigate the cost-efficiency of sampling in other regions of Africa and for other helminth infections, including schistosomiasis and lymphatic filariasis. This work is the subject of ongoing research.

Received November 25, 2009. Accepted for publication January 28, 2010.

**Acknowledgments:** We thank Charles Mwandawiro, Narcis Kabatereine, Christopher Milupi, and Faith Nchito for their contributions to the data collection and Jan Kolaczinski for constructive comments on an earlier draft of this paper.

**Financial support:** Hugh J. W. Sturrock is supported by a London School of Hygiene and Tropical Medicine Graduate Teaching Assistantship, Peter W. Gething is supported by a Wellcome Trust Senior Research Fellowship held by Dr. Simon Hay at the University of Oxford (#079091), and Simon Brooker is supported by a Research Career Development Fellowship from the Wellcome Trust (#081673).

**Authors' addresses:** Hugh J. W. Sturrock, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, E-mail: hugh.sturrock@lshtm.ac.uk. Peter W. Gething, Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, Oxford, United Kingdom, E-mail: peter.getting@zoo.ox.ac.uk. Archie C. A. Clements, School of Population Health, University of Queensland, Herston, Queensland, Australia; and Australian Centre for International and Tropical Health, Queensland Institute of Medical Research, Herston, Queensland, Australia, E-mail: a.clements@uq.edu.au. Simon Brooker, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Kenya Medical Research Institute-Wellcome Trust Research Programme, Nairobi, Kenya, E-mail: simon.brooker@lshtm.ac.uk.

## REFERENCES

- World Health Organization, 2006. *Neglected Tropical Diseases: Hidden Successes, Emerging Opportunities*. Available at: [http://whqlibdoc.who.int/hq/2006/WHO\\_CDS\\_NTD\\_2006.2\\_eng.pdf](http://whqlibdoc.who.int/hq/2006/WHO_CDS_NTD_2006.2_eng.pdf).
- World Health Organization, 2006. *Preventive Chemotherapy in Human Helminthiasis. Coordinated Use of Anthelmintic Drugs in Control Interventions: A Manual for Health Professionals and Programme Managers*. Geneva: World Health Organization.
- Baker MC, Mathieu E, Fleming FM, Deming M, King JD, Garba A, Koroma JB, Bockarie M, Kabore A, Sankara DP, Molyneux DH, 2010. Mapping, monitoring, and surveillance of neglected tropical diseases: towards a policy framework. *Lancet* 375: 231–238.
- Brooker S, Clements AC, Bundy DA, 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Adv Parasitol* 62: 221–261.
- Kurowski C, Wyss K, Abdulla S, Mills A, 2007. Scaling up priority health interventions in Tanzania: the human resources challenge. *Health Policy Plan* 22: 113–127.
- Mathieu E, Deming M, Lammie PJ, McLaughlin SI, Beach MJ, Domingue J, Addiss DG, 2003. Comparison of methods for estimating drug coverage for filariasis elimination, Leogane Commune, Haiti. *Trans R Soc Trop Med Hyg* 97: 501–505.
- Vlassoff C, Tanner M, 1992. The relevance of rapid assessment to health research and interventions. *Health Policy Plan* 7: 1–9.
- Anker M, 1991. Epidemiological and statistical methods for rapid health assessment: introduction. *World Health Stat Q* 44: 94–97.
- Brooker S, Kabatereine NB, Gyapong JO, Stothard JR, Utzinger J, 2009. Rapid mapping of schistosomiasis and other neglected tropical diseases in the context of integrated control programmes in Africa. *Parasitology* 136: 1707–1718.
- Lengeler C, Utzinger J, Tanner M, 2002. Questionnaires for rapid screening of schistosomiasis in sub-Saharan Africa. *Bull World Health Organ* 80: 235–242.
- Brooker S, Kabatereine NB, Myatt M, Russell Stothard J, Fenwick A, 2005. Rapid assessment of *Schistosoma mansoni*: the validity, applicability and cost-effectiveness of the lot quality assurance sampling method in Uganda. *Trop Med Int Health* 10: 647–658.
- Rabarijaona LP, Boisier P, Ravaoalimalala VE, Jeanne I, Roux JF, Jutand MA, Salamon R, 2003. Lot quality assurance sampling for screening communities hyperendemic for *Schistosoma mansoni*. *Trop Med Int Health* 8: 322–328.
- 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. *Bull World Health Organ* 81: 877–885.
- 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. *Trans R Soc Trop Med Hyg* 95: 267–269.
- World Health Organization, 2005. *Monitoring and Epidemiological Assessment of the Programme to Eliminate Lymphatic Filariasis at Implementation Unit Level*. Geneva: World Health Organization.
- Noma M, Nwoke BE, Nutall I, Tambala PA, Enyong P, Namsenmo A, Remme J, Amazigo UV, Kale OO, Seketeli A, 2002. Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). *Ann Trop Med Parasitol* 96 (Suppl 1): S29–S39.
- Takougang I, Meremikwu M, Wandji S, Yenshu EV, Aripko B, Lamle SB, Eka BL, Enyong P, Meli J, Kale O, Remme JH, 2002. Rapid assessment method for prevalence and intensity of *Loa loa* infection. *Bull World Health Organ* 80: 852–858.
- Brooker S, Kabatereine NB, Tukahebwa EM, Kazibwe F, 2004. Spatial analysis of the distribution of intestinal nematode infections in Uganda. *Epidemiol Infect* 132: 1065–1071.
- World Health Organization, 1991. *Basic Laboratory Methods in Medical Parasitology*. Geneva: World Health Organization.
- Kabatereine NB, Tukahebwa EM, Kazibwe F, Twa-Twa JM, Barenzi JF, Zaramba S, Stothard JR, Fenwick A, Brooker S, 2005. Soil-transmitted helminthiasis in Uganda: epidemiology and cost of control. *Trop Med Int Health* 10: 1187–1189.
- Clements AC, Lwambo NJ, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, Webster JP, Fenwick A, Brooker S, 2006. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop Med Int Health* 11: 490–503.
- Diggle PJ, Ribeiro PJ Jr, 2007. *Model-Based Geostatistics*. New York: Springer.
- Oliver MA, Muir KR, Webster R, Parkes SE, Cameron AH, Stevens MC, Mann JR, 1992. A geostatistical approach to the analysis of pattern in rare disease. *J Public Health Med* 14: 280–289.
- Cressie N, 2000. Geostatistical methods for mapping environmental exposures. Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. *Spatial Epidemiology, Methods and Applications*. Oxford: Oxford University Press, 185–204.
- Ribeiro PJ, Jr, Diggle PJ, 2001. *GeoR: A Package for Geostatistical Analysis*. R-NEWS 1. Available at: <http://cran.r-project.org/doc/Rnews>.
- Goovaerts P, 1997. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.
- Brooker S, Kabatereine NB, Smith JL, Mupfasoni D, Mwanje MT, Ndayishimiye O, Lwambo NJ, Mbotha D, Karanja P, Mwandawiro C, Muchiri E, Clements AC, Bundy DA, Snow RW, 2009. An updated atlas of human helminth infections: the example of east Africa. *Int J Health Geogr* 8: 42.
- Booth M, Bundy DA, 1995. Estimating the number of multiple-species geohelminth infections in human communities. *Parasitology* 111: 645–653.
- World Health Organization, 2006. *Preventative Chemotherapy in Human Helminthiasis*. Geneva: World Health Organization.



30. World Bank, 1993. *World Development Report 1993: Investing in Health*. New York: Oxford University Press.
31. Walker D, Kumaranayake L, 2002. Allowing for differential timing in cost analyses: discounting and annualization. *Health Policy Plan* 17: 112–118.
32. 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: e402.
33. Brooker S, Kabatereine NB, Fleming F, Devlin N, 2008. Cost and cost-effectiveness of nationwide school-based helminth control in Uganda: intra-country variation and effects of scaling-up. *Health Policy Plan* 23: 24–35.
34. Mabaso ML, Appleton CC, Hughes JC, Gouws E, 2003. The effect of soil type and climate on hookworm (*Necator americanus*) distribution in KwaZulu-Natal, South Africa. *Trop Med Int Health* 8: 722–727.
35. Saathoff E, Olsen A, Sharp B, Kvalsvig JD, Appleton CC, Kleinschmidt I, 2005. Ecologic covariates of hookworm infection and reinfection in rural Kwazulu-natal/south Africa: a geographic information system-based study. *Am J Trop Med Hyg* 72: 384–391.
36. Brooker S, 2007. Spatial epidemiology of human schistosomiasis in Africa: risk models, transmission dynamics and control. *Trans R Soc Trop Med Hyg* 101: 1–8.
37. Srividya A, Michael E, Palaniyandi M, Pani SP, Das PK, 2002. A geostatistical analysis of the geographic distribution of lymphatic filariasis prevalence in southern India. *Am J Trop Med Hyg* 67: 480–489.
38. Alexander ND, Moyeed RA, Hyun PJ, Dimber ZB, Bockarie MJ, Stander J, Grenfell BT, Kazura JW, Alpers MP, 2003. Spatial variation of *Anopheles*-transmitted *Wuchereria bancrofti* and *Plasmodium falciparum* infection densities in Papua New Guinea. *Filaria J* 2: 14.
39. Shope RE, 1999. Factors influencing the geographic distribution and incidence of tropical infectious diseases. Guerrant RL, Walker DH, Weller PF, eds. *Tropical Infectious Diseases. Principles, Pathogens and Practice*. London: Churchill Livingstone, 16–21.
40. Gyapong JO, Remme JH, 2001. The use of grid sampling methodology for rapid assessment of the distribution of *Bancroftian filariasis*. *Trans R Soc Trop Med Hyg* 95: 681–686.
41. Ansell J, Guyatt HL, 2002. Comparative cost-effectiveness of diagnostic tests for urinary schistosomiasis and the implications for school health programmes. *Ann Trop Med Parasitol* 96: 145–153.
42. 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. *Trop Med Int Health* 5: 192–202.
43. Muennig P, Pallin D, Sell RL, Chan MS, 1999. The cost effectiveness of strategies for the treatment of intestinal parasites in immigrants. *N Engl J Med* 340: 773–779.
44. Williams B, Gopi PG, Borgdorff MW, Yamada N, Dye C, 2008. The design effect and cluster samples: optimising tuberculosis prevalence surveys. *Int J Tuberc Lung Dis* 12: 1110–1115.
45. Jovani R, Tella JL, 2006. Parasite prevalence and sample size: misconceptions and solutions. *Trends Parasitol* 22: 214–218.
46. Clements AC, Bosque-Oliva E, Sacko M, Landoure A, Dembele R, Traore M, Coulibaly G, Gabrielli AF, Fenwick A, Brooker S, 2009. A comparative study of the spatial distribution of schistosomiasis in Mali in 1984–1989 and 2004–2006. *PLoS Negl Trop Dis* 3: e431.