ECOLOGIC COVARIATES OF HOOKWORM INFECTION AND REINFECTION IN RURAL KWAZULU-NATAL/SOUTH AFRICA: A GEOGRAPHIC INFORMATION SYSTEM–BASED STUDY

ELMAR SAATHOFF, ANNETTE OLSEN, BRIAN SHARP, JANE D. KVALSVIG, CHRIS C. APPLETON, AND IMMO KLEINSCHMIDT

Harvard School of Public Health, Boston, Massachusetts; Danish Bilharziasis Laboratory, Charlottenlund, Denmark; Malaria Lead Programme of the Medical Research Council, Durban, South Africa; Child, Youth and Family Development, Human Sciences Research Council, Dalbridge, South Africa; School of Life and Environmental Sciences, University of KwaZulu-Natal, Durban, South Africa

Abstract. Information on hookworm infection and re-infection in a cohort of primary school children and interview data on their socioeconomic background and behavior were combined with environmental data using a geographic information system (GIS). Multivariate models served to explore the covariation of environmental and infection patterns adjusted for possible confounders. Our aim was to identify environmental factors that might serve to predict infection and thus guide control efforts when epidemiologic information is insufficient. Furthermore, we wanted to establish whether soil type has a genuine influence on hookworm infection. Prevalence maps and spatial statistics showed considerable spatial clustering of infection in the small (~28 × 16 km) study area. The multivariate logistic regression models showed strong positive associations of infection at baseline (baseline prevalence = 83.2%) with settlement density (odds ratio [OR] = 1.24, 95% confidence interval [CI] = 1.10–1.38) and vegetation density (OR = 1.66, 95% CI = 1.25–2.22) and a strong negative association with the clay content of the soil (OR = 0.67, 95% CI = 0.62–0.73). Similar but weaker correlations were found after re-infection. Socioeconomic status and behavior did not seem to confound these associations. Spatial analysis of the model residuals suggested that because the models accounted for most of the spatial pattern, the model standard errors were not affected by spatial clustering. Our study shows that the pattern of hookworm infection is strongly influenced by several environmental factors. The GIS-aided prediction of areas in need of treatment is therefore a promising tool to guide control efforts when epidemiologic information is insufficient.

INTRODUCTION

Hookworm infection is common in tropical and subtropical parts of the world and especially prevalent in developing countries. Estimates suggest that the number of infections is more than 700 million worldwide with approximately 200 million occurring in sub Saharan-Africa.[1]

Geographic information systems (GISs) have gained importance for research into and prediction of the occurrence of parasitic diseases during the last two decades.[2] The rationale behind this is that parasitic diseases where the parasite life-cycle either entails a free-living stage or an animal intermediate host are more sensitive to environmental influences than diseases with direct person to person transmission. A GIS provides a convenient way to integrate large spatially referenced datasets concerning various aspects of geographic entities into one spatial database that can be used for research into disease etiology and for disease prediction. This integration of datasets is particularly important since it avoids the need for lengthy and costly data collection while providing a means for spatial prediction that can direct control efforts to the worst affected areas.[3]

A GIS has been used to study and predict the occurrence of malaria, trypanosomiasis, schistosomiasis, and tick-borne and numerous other parasitic diseases.[4–10] However, there are only few published reports where a GIS is used to investigate the association of environmental conditions with intestinal helminth infections.[11]

It is generally believed, that hookworms thrive in areas with sandy soils.[11–13] However, evidence in support of this hypothesis either originates from experiments that were concerned with larval survival and did not examine transmission,[14,15] or from ecologic studies[16] that were unable to ac-

count for confounders,[17–20] such as climatic, socioeconomic, and other factors.

The objective of this study was to explore associations of hookworm infection with environmental factors to further investigate the environmental epidemiology of this helminthiasis and to identify possible disease predictors. More specifically, we tried to establish whether hookworm infection is actually influenced by the sand content of the soil. The small spatial scale of the study allowed for collection of socioeconomic and behavioral data at an individual level and avoided bias caused by large-scale climatic differences. This enabled us to effectively control for confounders.

MATERIALS AND METHODS

Study area, population, and parasitologic data collection. The study was conducted between March 1998 and July 1999 in the Ingwavuma District in Maputaland, a rural and relatively undeveloped part of northern KwaZulu-Natal, South Africa (Figure 1). This area was selected because an earlier study had shown that helminth prevalences were high and exhibited strong spatial pattern.[21] The terrain comprises a flat area with very sandy soils east of the floodplain of the perennial Pongola River. The more elevated parts to the west of the river show a higher variation in soil type. The floodplain itself is dominated by loamy alluvial deposits and parts of it are used for subsistence agriculture. The floodplain is not inhabited because of regular flooding during the summer. Climate in the area is tropical to subtropical with a hot and humid summer (November–February) and a cooler and dry winter (June–August). Table 1 presents some of the environmental characteristics of the study area as they were found in the surroundings of the children's homes. A detailed map of
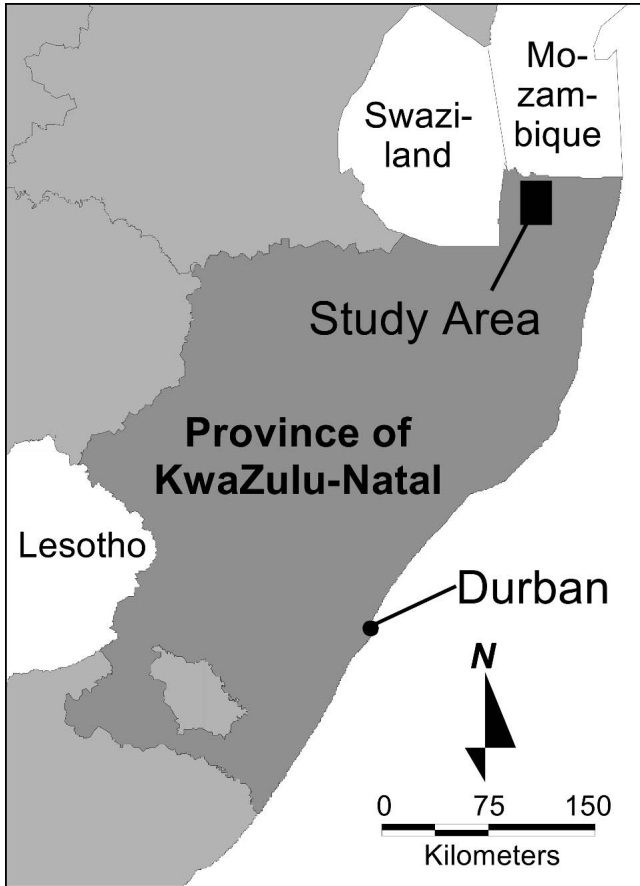
FIGURE 1. Location of the study area in northern KwaZulu-Natal.

TABLE 2
Characteristics of the study population at the time of the baseline survey (n = 825)

Characteristics	% with characteristic or median
Living in western part of study area	65.6%
Median age in years	10.8
(interquartile range of age)	(9.7–12.3)
Female	56.7%
Mother present	95.8%
Mother literate	52.6%
Mother employed	18.7%
Father present	87.5%
Father literate	70.4%
Father employed	59.5%

treatment (April–May 1999). More detailed information concerning the geohelminth infection status of the population at baseline, impact of treatment, and re-infection is reported elsewhere.[23]

In each parasitologic survey, schools were visited at least three times to include children who were absent or unable to deliver a stool specimen on the first occasion. On our visits, pupils were provided with the necessary sampling equipment and asked to provide a stool specimen. Apart from the pre-treatment baseline survey, where time limitations only allowed for the collection of one specimen, we tried to obtain two stool specimens per pupil. The few pupils who only provided one specimen (less than 10% in both post-treatment surveys) were nevertheless included in the analysis to minimize selection bias.

Specimens were kept cool prior to preparation of the slides. Duplicate 50-mg Kato-Katz cellophane thick smears were prepared from every fecal sample and were examined for hookworm eggs within one hour after preparation.[24] Diarrheal specimens and slides that were too dark were not examined; instead the respective pupils were asked to provide another specimen. Infection intensities are expressed as eggs per gram (epg) of feces. For individuals, these were calculated as the arithmetic mean of the number of eggs per thick smear multiplied by 20. The geometric mean of the individual values was used to characterize infection intensity on a population level. Children who still passed hookworm eggs after the two rounds of treatment were not included in the analyses of re-infection.

Ethical clearance was obtained from the Ethics Committee of the Faculty of Medicine of the University of Natal/Durban and the study was also reviewed and approved by the Central Medical Ethics Committee in Denmark. Before the onset of the study, information meetings were held with the staff and parents of the schools in the study. At these meetings, informed consent was obtained from the parents. Informed consent from the children was obtained directly before the first specimen collection.

Socioeconomic, behavioral, and environmental data. Interviews with the children were conducted during June and July 1999, immediately after the last parasitologic survey. The interviewers were not aware of the actual or former infection status of the children. During the interviews, children were asked about behavior and knowledge relevant to worm infections and about markers of socioeconomic status.

Environmental data were obtained from various sources:

the study area (including soil-types) and a climate chart are available on the internet.[22]

The study population was limited to pupils attending grade 3 in all ten schools in the study area. Characteristics of the children who participated in the baseline survey are presented in Table 2. The parasitologic data presented here were collected during three surveys. A baseline survey in March 1998 was followed by two treatments (April–May 1998 and October 1998) with 400 mg of albendazole (Zentel®; GlaxoSmith-Kline Beecham, Brentford, United Kingdom) of all children irrespective of infection status. Another survey to detect failed treatment was conducted three weeks after the second treatment (October–November 1998). Re-infection was assessed during the last survey at 29 weeks after the second

TABLE 1
Environmental conditions in the surroundings of the children's homesteads (n = 716)

Environmental factor	Median	Interquartile range
Population density in persons/km^2	67	45–140
Clay content of soil, %	8.6	3.2–8.6
Annual average of NDVI*	0.456	0.439–0.468
Elevation in meters	60	49–82
Slope in degrees	0.89	0.63–2.10

* The normalized difference vegetation index (NDVI) is a unitless satellite-sensored measure of chlorophyll abundance that is used to quantify vegetation coverage. It can theoretically take on values between −1 and +1, but usually falls within a range from 0.0 to 0.8.

Information on population density (persons/km^2) was obtained from the 1996 population census and settlement density information (as a map of homesteads in the area) was obtained from the South African Medical Research Council GIS malaria database for the area.[10,25] Because information about the sand content of the soil or its porosity was difficult to obtain, we used the soil clay content as a reverse proxy. This was estimated using a land type map of the area and the accompanying key that provide detailed information on soil composition.[26] Elevation, slope, and exposure to the sun were calculated using elevation contours from electronic versions of 1:50.000 topographic maps of the area and the respective surface analysis components of Idrisi32 GIS software.[27,28]

Vegetation density was characterized using data for the normalized difference vegetation index (NDVI) with a spatial resolution of 1 km. These data are distributed at no cost to the user by the Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey EROS Data Center (Sioux Falls, SD) (http://LPDAAC.usgs.gov).[29] It is obvious that data recorded closer to and during the study period might have better reflected the actual vegetation conditions. However, the alternatives would have been to either use 8-km resolution data, which seem too coarse for an area of approximately 28 × 16 km, or to buy higher resolution data for the respective period. Since we were concerned that our results should be of direct relevance to future studies, we decided to use the 1-km data because they are affordable, relatively easy to use, and available with worldwide coverage.

The basis for relating these data to individuals were the homestead positions of the children that we had previously determined using a global positioning system (GPS) receiver (GPS 12; Garmin, Olathe, KS). The environmental data were averaged for an area within a 1-km radius around the home of each child. This radius was chosen since a buffer of this size should represent the environmental conditions around the home relatively well.

Data analysis. Data were double-entered and corrected for data entry errors. Statistical analysis was carried out in Stata 7 for Windows (Stata Corp., College Station, TX) and Microsoft (Redmond, WA) Excel® was used to create the related figures. Prevalence maps were calculated and drawn using ArcView 3.2.[30,31]

Exploratory analysis showed that infection was clustered on several levels: children from the same home, same school and the same part of the study area (east or west of the Pongola River) were likely to be more similar to each other regarding hookworm infection when compared with children from elsewhere. In addition, we observed limited positive spatial autocorrelation, meaning that children who were living close to each other were also more similar with regard to hookworm infection.[32]

Clustering and spatial correlation of responses invalidates the statistical assumption of independence of observations, resulting in an under-estimation of standard errors and thus exaggerated statistical significance.[33–35] Several techniques have been proposed to overcome the problems of hierarchical clustering and spatial autocorrelation, but there is no single accepted way to do this for non-Gaussian data.[36–38]

To overcome the problem of clustering within homesteads (several participants shared their homes with one, two, or three other participants) dependent and independent variables were averaged for each home, and the homestead was made the unit of analysis. This reduced the effective sample size from 825 pupils to 716 participant homes for the baseline data and from 753 to 666 for the re-infection data. Binary disease status of the homestead for logistic regression purposes was coded as 0 if only uninfected participants lived there, and as 1 if at least one of the participating children was infected. Clustering on other levels was addressed during analysis.

Model identification. Data for hookworm infection at baseline and after re-infection were analyzed separately. Explanatory variables for both outcomes were selected systematically in three consecutive steps: first socioeconomic factors that were associated with hookworm infection and were therefore regarded as possible confounders were identified. Similarly, behavioral factors and in a third step environmental factors that were associated with hookworm infection were selected.

Stepwise forward logistic regression was used to select the best model in each case. *P* values for inclusion and exclusion of variables during the stepwise procedure were set to 0.10 and 0.11 respectively. Age, sex, and residence in the eastern or western part of the study area were retained in all models during the model identification process. Robust (i.e., "Huber-White" or "Sandwich") estimates of the standard errors were calculated to allow for clustering of responses within the same school.[39]

Resulting models. Single variable logistic regressions were carried out to assess unadjusted correlations of the identified predictor variables with the disease outcomes. Because we wanted to determine whether the results of the multivariate models were sensitive to the technique used to adjust for clustering of responses, we used two alternative methods to model within-cluster correlation in addition to the above described ordinary logistic model: 1) Generalized estimating equations (GEEs) logistic models with exchangeable correlation structure and with school as the unit of clustering;[40] and 2) negative binomial regression modeling with the untransformed egg count data (epg) as the outcome that, like the logistic models, also used robust standard errors to adjust for clustering within schools. Effect estimates of the negative binomial models were labeled egg count ratios by analogy to incidence rate ratios.

Spatial statistics. Spatial autocorrelation was assessed using two different approaches. Semi-variograms were drawn using GS+ software for the calculations and Microsoft Excel® 97 to produce the graphs. Because the semi-variogram approach does not provide estimates of statistical significance Moran's *I* and the related *P* values were calculated using the Stata 7 module spatgsa.[41,42]

RESULTS

The prevalence of hookworm infection at baseline was 83.2%, with boys being slightly more often infected than girls (86.5% versus 80.6%). Infected children had a geometric mean intensity of infection of 362 epg. Twenty-nine weeks after the second treatment, the prevalence in those children who had been egg negative after treatment was 33.9% with a geometric mean intensity of 118 epg. Again, the prevalence was higher in boys than in girls (36.3% versus 32.1%).

Figure 2 demonstrates spatial clustering of infection in the study area. Apart from the generally focal distribution, it
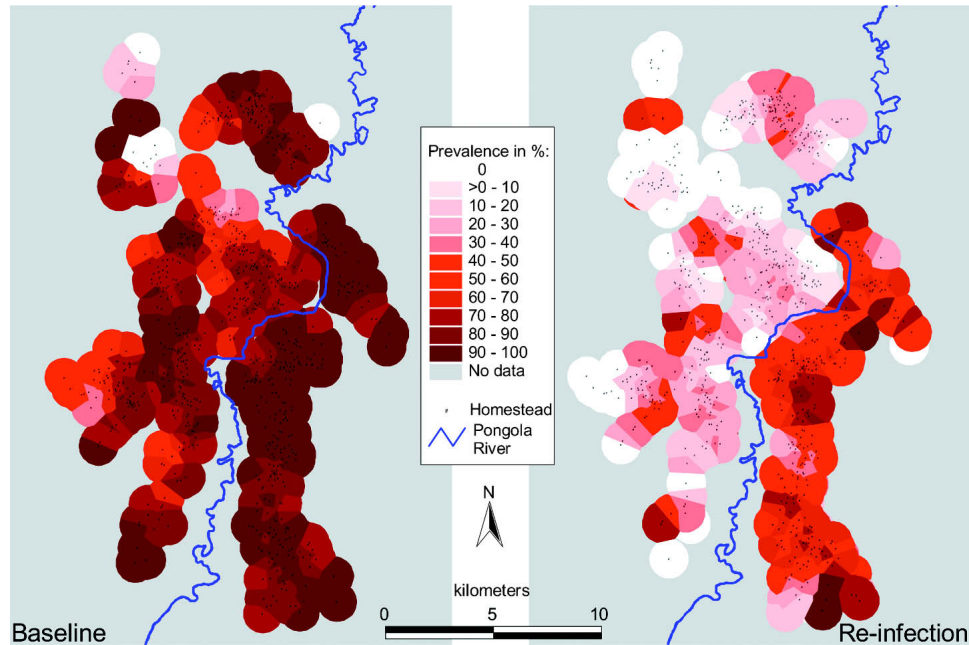
FIGURE 2.   Prevalence of hookworm infection at baseline and after 29 weeks of re-infection. Values were calculated for an area of 1 km radius around each homestead and assigned to Thiessen polygons drawn on the basis of homestead positions.[71]

shows that hookworm infection was much more prevalent in the eastern part of the study area than in the western part. Prevalence in the two areas was 94.3% versus 76.6% at baseline and 59.6% versus 20.8% after re-infection.

Age was a strong predictor of infection at baseline, but apparently played no role in re-infection and was therefore not included in the respective final models. The following socioeconomic and behavioral variables were considered for inclusion into the final models: the presence of certain assets in the household of the child (radio, refrigerator, television, western furniture, car, electricity), which were combined in a score; household ownership of cattle; home-gardening of food; the number of meals per week containing fish and containing meat; presence, employment status, and literacy of the mother and father; religion of the child; ownership and wearing frequency of shoes; sanitary habits (hand washing, soap usage, toilet availability and usage); cleanliness of fingernails during interview; and health knowledge about geohelminths and schistosomes which was again combined in a score. Because the focus of this paper is on environmental risk factors for hookworm infection, socioeconomic and behavioral factors will not be discussed in detail. However, those factors that were associated with infection in the study area are shown in Table 3 together with the environmental factors.

The multivariate models (Table 3) show that the clay content of the soil (3–43%), vegetation density, and population density are important predictors of infection, whereas elevation (40–162 meters above sea level), slope (0–4.3°), and exposure to the sun (a relative and unitless measure) do not seem to play an important role in the study area. These associations are stronger at baseline than after re-infection. Table 3 only shows the best fit model as identified in our analyses, but when we replaced July–September NDVI with NDVI measures for other periods (e.g., NDVI averages for every single month, for the whole year, for periods of two months, etc.), all were positively associated with hookworm

infection at baseline, and half of these associations were statistically significant at the 95% level of confidence. Results were quite similar when applying the same procedure to the re-infection data, but with a lower proportion of significant associations of NDVIs for the different periods. During this statistical examination, it became apparent that the longer term aggregated NDVI data (e.g., three-month, half year, and yearly NDVI averages) were stronger and more consistently associated with infection intensity than monthly and two-month data. However, a seasonal pattern for the association between NDVI and hookworm infection was not discernible.

The results of the different multivariate models agree relatively well with each other, but there are some differences when comparing them with the unadjusted results. Most obvious are these differences for the predictors settlement density (at baseline) and population density (after re-infection), where analysis showed that the unadjusted associations were confounded by vegetation density and clay content (at baseline) and by residence in the eastern or western part of the study area (after re-infection).

When comparing the ordinary and the GEE logistic model with the negative binomial model, it should be kept in mind that the former two report odds ratios, whereas the effect estimates of the negative binomial model are closely related to incidence rate ratios. In a high prevalence situation, they should therefore be expected to differ numerically.[43]

The semi-variogram (Figure 3) shows how similar or dissimilar observations are at different separation distances and thus characterizes spatial autocorrelation in the data. Every point in the graph represents the semi-variance of all possible pairs of observations that are separated by the same distance. To give an example: The first diamond from the left in the baseline variogram characterizes the variability of hookworm infection status within all pairs of children that lived between > 0 and 250 meters from each other, the next diamond stands for those who lived between 250 and 500 meters apart, etc.

Unadjusted and multivariate associations of hookworm infection with socioeconomic and environmental factors at baseline (n = 716) and after re-infection (n = 666)*

	Unadjusted†‡		Ordinary logistic‡		GEE logistic§		Negative binomial‡	
	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	ECR (95% CI)	P
Infection at baseline								
Part of study area, east relative to west	5.063 (2.55, 10.1)	< 0.001	1.85 (1.00, 3.41)	0.049	1.74 (0.98, 3.10)	0.059	1.06 (0.76, 1.47)	0.741
Male relative to female	1.494 (0.83, 2.69)	0.180	1.58 (1.00, 2.51)	0.051	1.58 (1.01, 2.49)	0.047	1.25 (1.00, 1.57)	0.054
Age per year	1.157 (0.94, 1.42)	0.160	1.20 (1.01, 1.43)	0.039	1.21 (1.02, 1.44)	0.030	1.06 (1.02, 1.10)	0.001
Mother employed relative to unemployed or absent	0.478 (0.29, 0.78)	0.003	0.45 (0.29, 0.69)	< 0.001	0.44 (0.29, 0.67)	< 0.001	0.72 (0.58, 0.91)	0.005
Settlement density per 10 homes/km²	1.092 (0.75, 1.60)	0.649	1.24 (1.10, 1.38)	< 0.001	1.21 (1.09, 1.35)	< 0.001	1.11 (0.99, 1.25)	0.070
Clay content of soil per 10% of total	0.599 (0.53, 0.68)	< 0.001	0.67 (0.62, 0.73)	< 0.001	0.67 (0.61, 0.73)	< 0.001	0.81 (0.77, 0.85)	< 0.001
NDVI, Jul–Sep¶ per decile	3.192 (2.20, 4.64)	< 0.001	1.66 (1.25, 2.22)	0.001	1.78 (1.45, 2.18)	< 0.001	1.49 (1.27, 1.74)	< 0.001
Re-infection								
Part of study area, east relative to west	5.64 (4.23, 7.51)	< 0.001	6.53 (3.53, 12.1)	< 0.001	6.86 (3.76, 12.5)	< 0.001	4.29 (2.90, 6.34)	< 0.001
Male relative to female	1.20 (0.86, 1.65)	0.282	1.30 (0.93, 1.83)	0.128	1.32 (0.94, 1.85)	0.106	1.29 (1.00, 1.66)	0.053
Meat eating per meat meal/week	0.92 (0.85, 0.99)	0.041	0.89 (0.79, 0.99)	0.038	0.89 (0.80, 0.99)	0.033	0.91 (0.85, 0.98)	0.011
Mother employed relative to unemployed or absent	0.64 (0.40, 1.02)	0.059	0.54 (0.38, 0.77)	0.001	0.54 (0.38, 0.77)	0.001	0.62 (0.45, 0.85)	0.003
Fingernails dirty relative to clean	1.19 (0.95, 1.51)	0.136	1.35 (1.05, 1.75)	0.021	1.36 (1.05, 1.75)	0.018	1.25 (1.02, 1.53)	0.035
Population density per 100 persons/km²	0.93 (0.71, 1.21)	0.586	1.26 (1.05, 1.52)	0.014	1.31 (1.10, 1.55)	0.002	1.21 (1.02, 1.44)	0.032
Clay content of soil per 10% of total	0.57 (0.46, 0.70)	< 0.001	0.81 (0.67, 0.98)	0.028	0.84 (0.71, 1.00)	0.052	0.80 (0.68, 0.93)	0.004
NDVI, Dec–May¶ per decile	1.12 (0.49, 2.56)	0.779	1.45 (1.01, 2.08)	0.046	1.55 (1.09, 2.20)	0.015	1.30 (0.99, 1.71)	0.062

* OR = odds ratio; ECR = egg count ratio; CI = confidence interval.
† Results of separate logistic regression models for each predictor variable.
‡ Adjusted for clustering within school and using robust variance estimates.
§ Generalized estimating equations (GEE) model with the following specifications: family = binomial, link = logistic, correlation = exchangeable; adjusted for clustering within schools and using robust variance estimates.
¶ Average of normalized difference vegetation index (NDVI) for the specified period according to 1-km resolution Advanced Very High Resolution Radiometer satellite imagery. One unit represents 10% of the range between the lowest and the highest values found in the NDVI data.

Values less than one indicate that the variability at this distance is lower than the overall variability and thus signify positive spatial autocorrelation whereas values above unity indicate the opposite.

The trend in semi-variance over short distances is especially interesting. The steep increase with distance in semi-variance for the raw baseline data indicates that the occurrence of infection was autocorrelated over distances of up to 4 km where semi-variance reaches a plateau. This means that children who lived within a distance of less than 4 km from each other were more similar to each other than to the other pupils. The much gentler increase in the re-infection semi-variances indicates a weaker pattern of autocorrelation.

When comparing semi-variograms of the raw data with those of the respective deviance residuals, it is obvious that the latter were much less autocorrelated. This indicates that the independent variables included in the model explained most of the spatial correlation in infection status.
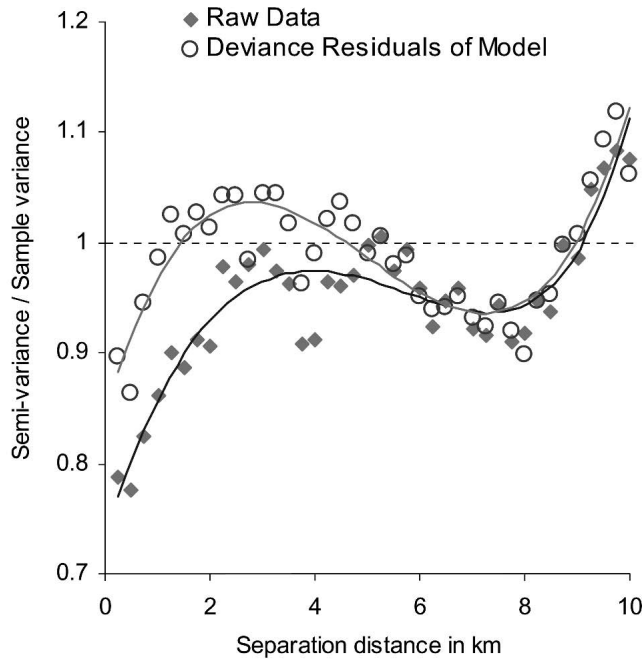
Table 4 is in agreement with this. It shows that spatial autocorrelation as indicated by Moran's *I* (which ranges from +1 = strong positive autocorrelation to −1 = strong negative autocorrelation) was significant in the raw data, but greatly

reduced in the model residuals. Nevertheless, the baseline data still display some significant residual autocorrelation for children living within 1,000 meters from each other, indicating unexplained residual correlation in infection rates.
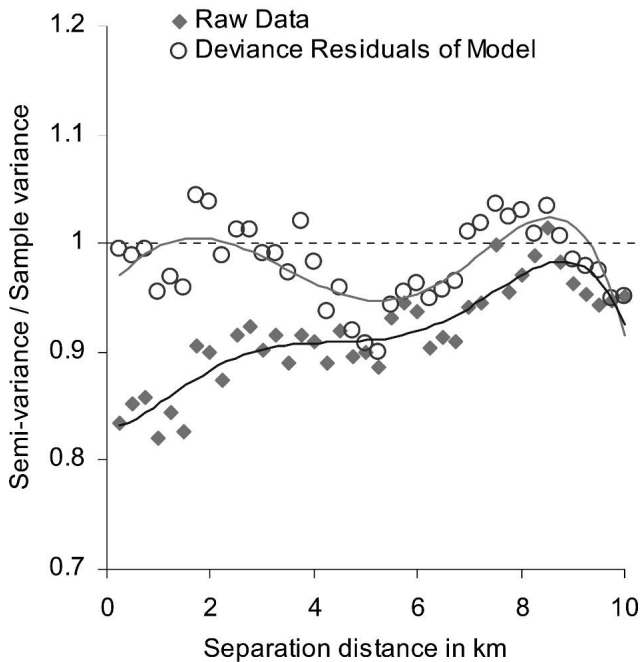
DISCUSSION

Our results demonstrate a strong negative association of hookworm infection with the clay content of the surface soil and a strong positive association with measures of vegetation and population density.

The clay content was included as an inverse proxy for soil porosity because data on particle size or sand content were not available. Soil texture is believed to be important because porous soils are well aerated and allow for vertical migration of the larvae, enabling them to actively avoid desiccation and solar radiation by moving deeper into the soil, and to avoid rising water levels by moving upwards. In contrast, clay soil seems to inhibit larval migration.[13,15,44–47] In accordance with this reasoning, a positive association of sandy soils with hookworm infection has been demonstrated in many studies.[17–19,48–50]

A. Baseline

B. Re-infection

FIGURE 3. Autocorrelation of binary hookworm infection status: omni-directional semi-variograms of the raw data and of deviance residuals of the logistic models for infection at **A**, baseline and **B**, for re-infection. Lag distance = 250 meters.

TABLE 4
Moran's I for autocorrelation in the raw hookworm infection data and in the deviance residuals of the logistic model

Separation distance (m)	Raw data		Residuals	
	I	P	I	P
Baseline				
0–1,000	0.153	< 0.001	0.036	0.018
0–2,000	0.121	< 0.001	0.010	0.189
0–3,000	0.089	< 0.001	0.001	0.696
0–4,000	0.077	< 0.001	0.004	0.341
Re-infection				
0–1,000	0.176	< 0.001	0.023	0.153
0–2,000	0.161	< 0.001	0.016	0.064
0–3,000	0.127	< 0.001	0.002	0.998
0–4,000	0.113	< 0.001	0.000	0.753

traviolet radiation. Associations of soil moisture and exposure to the sun with larval survival and infectivity have been demonstrated in numerous studies.[15,44,45,47,51–53] However, the only article we know of that demonstrates an association of NDVI with the occurrence of geohelminth infection does not consider hookworm.[54] Our study is therefore unique in demonstrating an association between hookworm and NDVI specifically.

The relationship of population density with hookworm infection seems also sensible because if sanitary facilities are lacking, fecal contamination of the soil increases with population, thus leading to elevated levels of transmission in disease-endemic areas.[55]

The lack of residual autocorrelation demonstrated in Figure 3 and Table 4 shows that our models accounted for the autocorrelation in the raw infection data. This is very important since it implies that the model standard errors are not underestimated due to the presence of spatial autocorrelation and therefore the calculated significance estimates are valid.[56]

As for the suspected socioeconomic and behavioral confounding, it may be worth noting that effect estimates and standard errors from models that did not include these factors but only contained the environmental covariates (not shown) were very similar to those of the full models shown in Table 3. This indicates that in the study area with its relatively homogenous population structure, socioeconomic and behavioral confounding seems to have been negligible. However, we believe that this conclusion cannot be extended to larger-scale studies with more diverse populations and locally clustered conditions.

It may be seen as a limitation of this study that information on socioeconomic status was obtained from the children and not from their parents. However, because most of our questions were concerned with relatively practical things (absence or presence of certain items in the household, number of cattle, literacy of mother and father, etc.) and thus quite straightforward, we believe that the information should nevertheless be valid. The only more complicated questions were concerned with the parental employment status. This information was collected by asking the child two questions about each parent: whether the mother or father was working, and also what the parent was doing to earn money. The information from the second question was used to validate the answer from the first question. In most cases, however, the two answers were in agreement.

The association of NDVI with hookworm infection that we found is also plausible. Soil moisture is an essential prerequisite for the development and infectivity of hookworm larvae and it is also an important determinant of plant cover. Plant cover provides shade and thus prevents desiccation of the top soil and shelters surface dwelling hookworm larvae from ul-

Our study shows that soil composition, vegetation, and

population density can strongly influence prevalence and intensity of hookworm infection and that they should therefore be considered when trying to develop predictive models and intervention strategies if there is insufficient epidemiologic evidence.

The NDVI and population density are especially promising for prediction because relatively detailed information with wide coverage for both factors is available in the public domain.[57,58] Soil composition is more problematic because detailed information is lacking for most developing countries where hookworm is most prevalent. Furthermore, comparisons of this measure between different countries pose problems because differences in survey methods, soil classification systems, etc. complicate a unified approach. However, with the advent of new remote sensing tools, this situation might soon change.[59,60]

The inclusion of socioeconomic factors that we believe are important determinants of large-scale variation is an even bigger challenge: detailed information is lacking, and the available information would be hard to generalize over country and cultural boundaries.

Our study did not assess the influence of climatic factors because we assumed that their variation in the study area would be small and because climatic data with the necessary high spatial resolution were not available. However, there is overwhelming evidence that rainfall and temperature are important determinants of the larger-scale geographic pattern of hookworm infection.[53,61–66]

Although elevation or altitude, which is closely related to the climatic factors, did not seem to play an important role in our small study area, it has been shown to be associated with the occurrence of hookworm infection in other studies on larger scales.[50,67,68] Like the NDVI and climate data, elevation data are available in the public domain at adequate resolution.[69,70]

In accordance with our expectations, this study shows that the pattern of hookworm infection is strongly influenced by several environmental factors. The GIS-aided prediction of areas in need of treatment is therefore a promising tool to guide control efforts when epidemiologic information is insufficient.

Authors' addresses: Elmar Saathoff, Harvard School of Public Health, 665 Huntington Avenue, Boston MA 02115, E-mail: elmarsaathoff@compuserve.de. Annette Olsen, Danish Bilharziasis Laboratory, Jaegersborg Allé 1 D, DK-2920, Charlottenlund, Denmark, E-mail: ao@bilharziasis.dk. Brian Sharp and Immo Kleinschmidt, Malaria Lead Programme of the Medical Research Council, PO Box 17120, Congella 4013, Durban, South Africa, E-mails: sharpb@mrc.ac.za and immo.kleinschmidt@mrc.ac.za. Jane D. Kvalsvig, Child, Youth and Family Development, Human Sciences Research Council, Private Bag X07, Dalbridge 4014, South Africa, E-mail: JKvalsvig@hsrc.ac.za. Chris C. Appleton, School of Life and Environmental Sciences, University of KwaZulu-Natal, Durban 4041, South Africa, E-mail: appleton@biology.und.ac.za.

REFERENCES

1. de Silva NR, Brooker S, Hotez PJ, Montresor A, Engels D, Savioli L, 2003. Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol 19:* 547–551.
2. Hay SI, Randolph SE, Rogers DJ, 2000. *Remote Sensing and Geographical Information Systems in Epidemiology*. San Diego, CA: Academic Press.
3. Mott KE, Nuttall I, Desjeux P, Cattand P, 1995. New geographical approaches to control of some parasitic zoonoses. *Bull World Health Organ 73:* 247–257.
4. Hay SI, Omumbo J, Craig M, Snow RW, 2000. Earth observation, geographic information systems and *Plasmodium falciparum* malaria in sub-Saharan Africa. *Adv Parasitol 47:* 174–215.
5. Craig MH, Snow RW, le Sueur D, 1999. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today 15:* 105–111.
6. Rogers DJ, 2000. Satellites, space, time and the African trypanosomiases. *Adv Parasitol 47:* 129–171.
7. Kristensen TK, Malone JB, McCarroll JC, 2001. Use of satellite remote sensing and geographic information systems to model the distribution and abundance of snail intermediate hosts in Africa: a preliminary model for *Biomphalaria pfeifferi* in Ethiopia. *Acta Trop 79:* 73–78.
8. Malone JB, Abdel-Rahman MS, El Bahy MM, Huh OK, Shafik M, Bavia M, 1997. Geographic information systems and the distribution of *Schistosoma mansoni* in the Nile Delta. *Parasitology Today 13:* 112–119.
9. Randolph SE, 2000. Ticks and tick-borne disease systems in space and from space. *Adv Parasitol 47:* 217–243.
10. Kleinschmidt I, Sharp BL, Clarke GP, Curtis B, Fraser C, 2001. Use of generalized linear mixed models in the spatial analysis of small - area malaria incidence rates in KwaZulu Natal, South Africa. *Am J Epidemiol 153:* 1213–1221.
11. Brooker S, Michael E, 2000. The potential of geographical information systems and remote sensing in the epidemiology and control of human helminth infections. *Adv Parasitol 47:* 245–288.
12. Beaver PC, 1952. Observation on the epidemiology of ascariasis in a region of high hookworm endemicity. *J Parasitol 38:* 445–453.
13. WHO, 1964. *Soil Transmitted Helminths—Report of a WHO Expert Committee on Helminthiases*. Geneva: World Health Organization.
14. Augustine DL, 1922. Investigations on the control of hookworm disease. X. Experiments on the length of life of infective hookworm larvae in soils. *Am J Hyg 2:* 177–187.
15. Vinayak VK, Chitkara NL, Chhuttani PN, 1979. Soil dynamics of hookworm larvae. *Indian J Med Res 70:* 609–614.
16. Last JM, 1988. *A Dictionary of Epidemiology*. Oxford, United Kingdom: Oxford University Press.
17. Smith AJ, 1903. Uncinariasis in Texas. *Am J Med Sci 126:* 768–798.
18. Augustine DL, Smillie WG, 1926. The relation of the type of soils of Alabama to the distribution of hookworm disease. *Am J Hyg 6:* 36–62.
19. Hsieh HC, Kang BT, Stoll NR, Chen ER, Reber EW, Kuo M, 1971. Relation of hookworm prevalence to soil characteristics in Liberia. *Chin J Microbiol 4:* 132–156.
20. Mabaso ML, Appleton CC, Hughes JC, Gouws E, 2004. Hookworm (*Necator americanus*) transmission in inland areas of sandy soils in KwaZulu-Natal, South Africa. *Trop Med Int Health 9:* 471–476.
21. Schutte CHJ, Eriksson IM, Anderson CB, Lamprecht T, 1981. Intestinal parasitic infections in black scholars in northern KwaZulu. *S Afr Med J 60:* 137–141.
22. Saathoff E, Olsen A, Magnussen P, Kvalsvig JD, Becker W, Appleton CC, 2004. Patterns of *Schistosoma haematobium* infection, impact of praziquantel treatment and re-infection after treatment in a cohort of schoolchildren from rural KwaZuluNatal/South Africa. *BMC Infect Dis 4:* 40. http://www.biomedcentral.com/1471-2334/1474/1427.
23. Saathoff E, Olsen A, Kvalsvig JD, Appleton CC, 2004. Patterns of geohelminth infection, impact of albendazole treatment and re-infection after treatment in schoolchildren from rural Kwa-

Zulu-Natal/South-Africa. *BMC Infect Dis 4:* 27. http://www.biomedcentral.com/1471-2334/1474/1427.

24. WHO, 1991. *Basic Laboratory Methods in Medical Parasitology.* Geneva: World Health Organization.

25. Statistics South Africa, 1999. *Population Census 1996—Enumerator Area Database.* Pretoria, South Africa: Statistics South Africa.

26. Land Type Survey Staff, 1986. *Land Types of the Map 2632 Mkuze.* Pretoria, South Africa: Department of Agriculture and Water Supply.

27. Chief Directorate Surveys and Mapping, 1980. *Maps 2632CC, 2632CD, 2732AA, 2732AB.* Mowbray, South Africa: Department of Land Affairs.

28. The Idrisi Project, 2002. *Idrisi32 GIS Software, Release I32.22.* Worcester, MA: Clark Laboratories.

29. Eidenshink JC, Faundeen JL, 1994. The 1 km AVHRR global land data set: first stages in implementation. *Int J Remote Sensing 15:* 3443–3462.

30. Stata Corp, 2001. *Stata Statistical Software: Release 7.0.* College Station, TX: Stata Corporation.

31. ESRI, 1999. *ArcView GIS 3.2.* Redlands, CA: Environmental Systems Research Institute, Inc.

32. Cliff AD, Ord JK, 1973. *Spatial Autocorrelation.* London: Pion.

33. "Student" (W. S. Gosset), 1914. The elimination of spurious correlation due to position in time or space. *Biometrika 10:* 179–180.

34. Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR, 1996. A survey of methods for analyzing clustered binary response data. *Int Stat Rev 64:* 89–118.

35. Gould P, 1970. Is statistix inferens the geographical name for a wild goose? *Econ Geogr 46:* 439–448.

36. Bryk AS, Raudenbush SW, 1992. *Hierarchical Linear Models.* London: Sage Publications.

37. Goldstein H, 1998. *Multilevel Statistical Models.* London: Edward Arnolds.

38. Robinson TP, 2000. Spatial statistics and geographical information systems in epidemiology and public health. *Adv Parasitol 47:* 82–128.

39. Stata Corp, 2001. *Stata users guide. Release 7.0.* College Station, Texas: Stata Corporation.

40. Liang K, Zeger S, 1986. Longitudinal data analysis using generalized linear models. *Biometrika 73:* 13–22.

41. Gamma Design Software, 1998. *GS+, 3.10.2 Beta.* Plainwell, MI: Gamma Design Software.

42. Pisati M, 2001. Tools for spatial data analysis. *Stata Tech Bull 60:* 21–37.

43. Rothman KJ, Greenland S, 1998. *Modern Epidemiology.* Philadelphia: Lippincott-Raven.

44. Augustine DL, 1922. Investigations on the control of hookworm disease. VIII. Experiments on the migration of hookworm larvae in soils. *Am J Hyg 2:* 162–187.

45. Payne FK, 1922. Investigations on the control of hookworm disease. XI. Vertical migration of infective larvae in the soil (preliminary report). *Am J Hyg 2:* 254–263.

46. Beaver PC, 1975. Biology of soil-transmitted helminths: the massive infection. *Health Lab Sci 12:* 116–125.

47. Udonsi JK, Nwosu AB, Anya AO, 1980. *Necator americanus:* population structure, distribution, and fluctuations in population densities of infective larvae in contaminated farmlands. *Z Parasitenkd 63:* 251–259.

48. Boerner MH, 1915. Presence, prevalence and geographic distribution of hookworm infection in Texas. *Texas State J Med 10:* 377–378.

49. Otto GF, Cort WW, Keller AE, 1931. Environmental studies of families in Tennessee infested with *Ascaris, Trichuris* and hookworm. *Am J Hyg 14:* 156–193.

50. 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.

51. Sturrock RF, 1967. Hookworm studies in Tanganyika (Tanzania): the results of a series of surveys on a group of primary school-children and observations on the survival of hookworm infective larvae exposed to simulated field conditions. *East Afr Med J 44:* 142–149.

52. Lilley B, Lammie P, Dickerson J, Eberhard M, 1997. An increase in hookworm infection temporally associated with ecologic change. *Emerg Infect Dis 3:* 391–393.

53. Udonsi JK, Atata G, 1987. *Necator americanus:* temperature, pH, light, and larval development, longevity, and desiccation tolerance. *Exp Parasitol 63:* 136–142.

54. Brooker S, Hay SI, Tchuem Tchuenté L, Ratard RC, 2002. Using NOAA-AVHRR data to model human helminth distributions in planning disease control in Cameroon, West Africa. *Photogrammetric Engineering Remote Sensing 68:* 175–179.

55. Schad GA, Nawalinski TA, Kochar V, 1983. Human ecology and the distribution and abundance of hookworm populations. Croll NA, Cross JH, eds. *Human Ecology and Infectious Disease.* New York: Academic Press, 187–223.

56. Cliff AD, Ord JK, 1981. *Spatial Processes—Models & Applications.* London: Pion.

57. Deichman U, 1997. *African Population Database.* Sioux Falls, SD: United Nations Environment Programme Global Resource Information Database. http://grid2.cr.usgs.gov/globalpop/africa/.

58. Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA, 2000. LandScan: A global population database for estimating populations at risk. *Photogrammetric Engin Remote Sensing 66:* 849–857.

59. Huh OK, Malone JB, 2001. New tools: potential medical applications of data from new and old environmental satellites. *Acta Trop 79:* 35–47.

60. Wood BL, Beck LR, Lobitz BM, Bobo MR, 2000. Education, outreach and the future of remote sensing in human health. *Adv Parasitol 47:* 331–344.

61. Cort WW, Schapiro L, Riley WA, Stoll NR, 1929. A study of the influence of the rainy season on the level of helminth infestations in a Panama village. *Am J Hyg 10:* 626–634.

62. Diesfeld HJ, 1970. Correlation between hookworm findings and climate as indicated by the temperature humidity relation in Kenya. *Z Tropenmed Parasitol 21:* 84–92.

63. Miller TA, 1970. Studies on the incidence of hookworm infection in east Africa. *East Afr Med J 47:* 354–363.

64. Nwosu AB, Anya AO, 1980. Seasonality in human hookworm infection in an endemic area of Nigeria, and its relationship to rainfall. *Trop Med Parasitol 31:* 201–208.

65. Hall A, Latham MC, Crompton DW, Stephenson LS, Wolgemuth JC, 1982. Intestinal parasitic infections of men in four regions of rural Kenya. *Trans R Soc Trop Med Hyg 76:* 728–733.

66. Appleton CC, Maurihungirire M, Gouws E, 1999. The distribution of helminth infections along the coastal plain of KwaZulu-Natal province, South Africa. *Ann Trop Med Parasitol 93:* 859–868.

67. Jemaneh L, Tedla S, 1984. The distribution of *Necator americanus* and *Ancylostoma duodenale* in school populations, Gojam and Gondar Administrative Regions. *Ethiop Med J 22:* 87–92.

68. Appleton CC, Gouws E, 1996. The distribution of common intestinal nematodes along an altitudinal transect in KwaZulu-Natal, South Africa. *Ann Trop Med Parasitol 90:* 181–188.

69. EROS data center, 1996. *GTOPO30: Global Digital Elevation Model.* Reston, VA: U.S. Geological Survey (USGS).

70. Hutchinson MF, Nix HA, McMahon JP, Ord KD, eds., 1995. *Topographic and Climate Database for Africa.* Canberra, Australia: Centre for Resource and Environmental Studies, The Australian National University.

71. Ripley BD, 1981. *Spatial Statistics.* New York: John Wiley & Sons.