

Predicting schistosome transmission in rural Uganda using water contact data from wearable GPS devices

Received: 27 June 2025

Accepted: 28 March 2026

Published online: 22 May 2026

 Check for updates

Fabian Reitzug¹, Melissa A. Iacovidou¹, Narcis B. Kabatereine², Betty Nabatte², Max T. Eyre³, Benjamin Tinkitina², Anatol M. Byaruhanga², Renaud Lambiotte⁴ & Goylette F. Chami¹✉

Granular spatial data and models of human behaviour are currently lacking for schistosomiasis. We collected ten days of wearable GPS logger data from 452 individuals in rural Uganda to model water contact as a proxy indicator of usage at 69 georeferenced open-water sites and 32 public taps or boreholes. Among participants, 63.9 and 33.2% visited at least one water site and tap or borehole, respectively. Exponential spatial decay models accurately predicted site-specific open-water contact (area under the receiver operating characteristic curve = 0.87) and tap or borehole usage (0.92). There was no evidence of tap or borehole usage influencing open-water contact. Incorporating mobility terciles did not improve simple spatial decay models. Integrating spatial decay-based estimates of open-water-site usage into an individual-based transmission model produced realistic estimates of one-year *Schistosoma mansoni* reinfection and provided a ranking of water sites contributing to transmission. Our spatial decay models offer scalable tools for focal interventions for schistosomiasis.

Schistosomiasis is caused by a waterborne parasitic flatworm and estimated to currently be present in over 250 million people globally, with most cases in rural sub-Saharan Africa¹. In endemic areas, households rely on unprotected, open-water sites due to the lack of safe water sources such as taps, boreholes or piped water². Freshwater bodies, including rivers, streams and lakes, are used for swimming, drinking water collection and laundry, with each activity conferring different risks of infection and reinfection through varying cercarial exposure³. Mass drug administration (MDA) with praziquantel alone is recognized by the World Health Organization (WHO) and the WHO Regional Office for Africa as insufficient for reducing schistosome transmission^{4,5}. Focal interventions to complement MDA are needed, yet granular spatial data are scarce and models that incorporate such information for targeting are unavailable.

Our current understanding of focality in human contact with open-water sites is limited by conventional assumptions that assign

households to sites. Despite open-water contact being heterogeneous within villages and households^{6,7}, it is commonly assumed that people use only the open-water site closest to their household^{2,8} or only those within their village⁹. Beyond this, several studies have suggested a direct correspondence between human mobility and open-water contact^{10–12}, but it remains unclear how mobility shapes spatial patterns of site usage and whether more realistic assignment rules can be formulated beyond simply nearest distance alone.

The main focal intervention complementary to MDA and endorsed by the 2022 WHO schistosomiasis control guidelines¹³ is the provision of safe water, sanitation and hygiene (WASH). Evidence for WASH remains mixed; a meta-analysis of 39 observational studies suggested that access to safe drinking water was associated with 0.53 lower odds of schistosome infection¹⁴. However, the study authors noted inconsistent WASH definitions across studies and did not focus on reinfection, limiting the ability of the analysis to untangle WASH influences from

¹Big Data Institute, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom. ²Division of Vector Borne and Neglected Tropical Diseases, Ministry of Health, Kampala, Uganda. ³London School of Hygiene and Tropical Medicine, London, United Kingdom. ⁴Mathematical Institute, University of Oxford, Oxford, United Kingdom. ✉e-mail: goylette.chami@ndph.ox.ac.uk

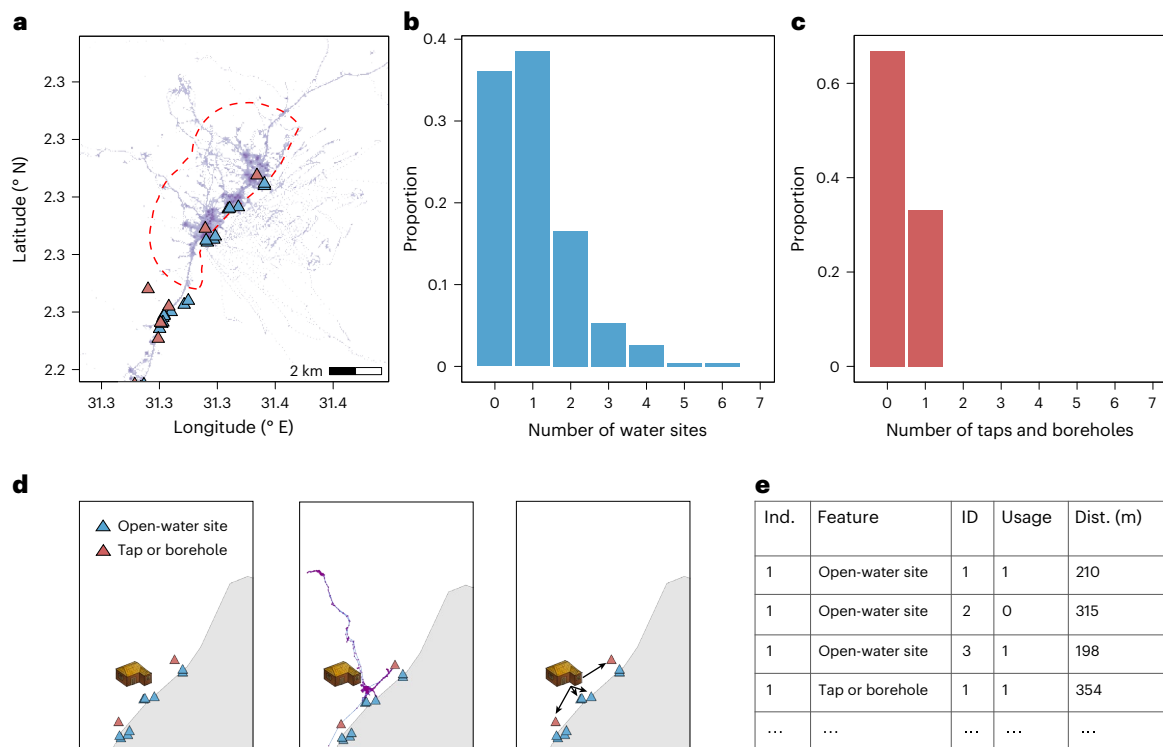


Fig. 1 | GPS logger data and processing. **a**, GPS logger data traces in Pakwach, aggregated across the study sample. GPS locations and movement are represented as purple lines. The blue triangles denote open-water sites mapped by the study team. The red triangles denote taps and boreholes. The red dashed line shows the study area. **b**, Number of distinct open-water sites visited by the study participants. **c**, Number of distinct taps and boreholes visited by the study participants. **d**, Processing steps to derive open-water-site and tap or borehole usage data from GPS data and household and site location data. **e**, Schematic of

the data frame resulting from the processing, which was used to fit spatial decay models. The columns represent the unique individual identifier (Ind.), type of water feature (open-water site or tap or borehole), water feature ID, whether the individual had any contact with the feature during the observation period (1 = yes and 0 = no) and Euclidean distance (Dist.) to the feature from the household location of the individual. Uganda waterbodies shapefile from World Resources Institute (<https://www.wri.org/>)⁷⁰. House icon adapted from Firkin/Openclipart.

other confounders, such as natural behavioural tendencies or participant characteristics, including socioeconomic status¹⁴. In contrast, a cluster randomized trial in Zanzibar reported no additional benefit of behaviour change (the provision of urinals and health education) and WASH provision (safe washing platforms), when combined with biannual MDA, compared with MDA alone¹⁵. As detailed measures of open-water and WASH usage were not collected, it remains challenging to determine why the intervention had no added benefits to biannual MDA. Most existing studies on water contact and WASH rely on self-reported information or household distance to sites and taps^{2,8,16–18} and therefore lack the objective, spatially granular information needed to characterize fine-scale water usage patterns and quantify the impact of WASH on water contact and (re)infection¹⁹.

Individual-based models (IBMs) have been developed to progress towards more focal interventions that account for heterogeneity in human behaviour. However, available IBMs for schistosome transmission, such as SCHISTOX²⁰ and SchiSTOP²¹, typically use average, age-group-specific water contact rates without spatial structure, despite demonstrated heterogeneity by waterbody distance, gender and occupation^{2,22,23} and evidence that such heterogeneity drives variation in infection prevalence²⁴. Existing IBMs therefore lack integration of human contact events at specific open-water sites with site-level environmental data, limiting their utility for targeting interventions to sites that sustain transmission.

To bridge these data and modelling gaps, we collected wearable global positioning system (GPS) logger data on 452 children and adults in the Pakwach and Buliisa districts of western Uganda and the Mayuge District in eastern Uganda, nested within the SchistoTrack cohort²⁵, where *Schistosoma mansoni* is highly endemic (prevalence = 43%)².

We studied open-water and tap or borehole usage over a ten-day period by combining GPS logger data with exhaustive mapping of all water sites used by the study communities, including those outside the study villages. The main aim was to model open-water contact and tap or borehole usage with fine-scale spatial data to develop assumptions about which individuals have contact with which open-water sites, and to validate these assumptions using spatially explicit IBMs of schistosome reinfection. A secondary aim was to evaluate whether models that incorporate WASH usage outperform models using spatial information alone in predicting human water contact at open-water sites.

Results

Patterns of open-water contact and tap or borehole usage

We derived indicators of open-water contact and tap or borehole usage, as shown in Fig. 1. Within our study catchment, 143 mapped open-water sites (including adjacent sites within larger waterbodies) and 63 taps or boreholes (including distinct facilities located next to each other) were clustered into 69 distinct open-water sites and 32 tap or borehole locations. Open-water site usage or contact was defined as at least one GPS-inferred contact with mapped, shallow open-water sites (<2 m maximum depth), as a proxy for human schistosome exposure at locations with the highest parasite density^{2,26}. Tap or borehole usage was defined as proximity (<20 m) to public taps or boreholes, comparable to buffer sizes used in other GPS logger studies (20–30 m)^{27,28}.

A total of 8,249 open-water contact events were observed across the three districts. Overall, 63.9% (289/452) of participants visited at least one open-water site; the median daily duration was 2 min (interquartile range (IQR) = 0–12.3 min) across all participants and 9 min (IQR = 3–22 min) among the 289 individuals with any contact, who

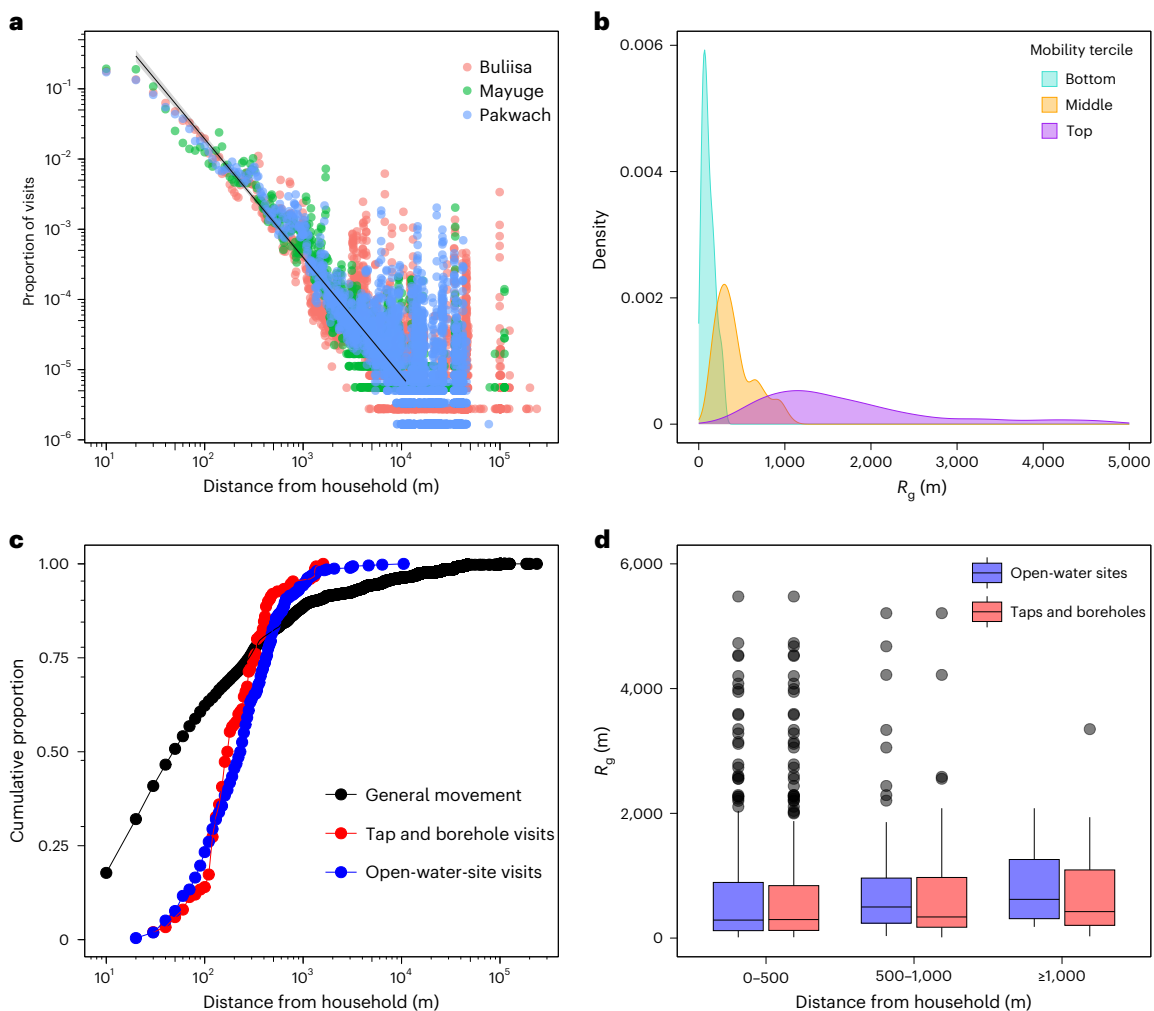


Fig. 2 | Association between mobility and usage of open-water sites and taps or boreholes. **a**, Proportion of visited locations over distance from the household. The black line represents a truncated power-law distribution fitted between 1.3×10^1 and 1.1×10^4 m with a 95% CI (shaded area). Limits were determined using the `powerLaw` R package. **b**, Distribution of radii of gyration (R_g ; equation (1)) by district-specific mobility tertile. R_g measures human mobility relative to the distance from the household. **c**, Cumulative proportions of visits to

open-water sites or tap and borehole locations, as well as all GPS locations (general movement), plotted against distance from the household. **d**, Relationship between R_g and household distance to the nearest open-water site or tap or borehole. Central horizontal lines and box limits represent median values and the IQR, respectively. The whiskers extend to the most extreme data points within $1.5 \times$ the IQR of each box. Outliers are plotted as individual points.

typically visited one distinct site (IQR = 1–2). The proportion visiting at least one open-water site varied substantially by district ($\chi^2 = 75.9$; d.f. = 2; $P < 0.01$), with 52.5% (96/183) in Pakwach, 89.6% (147/164) in Buliisa and 43.8% (46/105) in Mayuge. Diurnal patterns also differed by district; in Pakwach, open-water contact peaked at 08:00, whereas in Buliisa and Mayuge it peaked around 19:00 (Supplementary Fig. 1a).

We observed 1,181 tap or borehole usage events, 85.7% fewer than open-water contact events. Overall, 33.2% (150/452) of participants visited at least one tap or borehole. The proportion of participants using at least one tap or borehole was 49.4% (81/164) in Buliisa, 25.1% (46/183) in Pakwach and 21.9% (23/105) in Mayuge. The median daily duration of tap or borehole usage was 0 min (IQR = 0–1 min) across all participants. Among users, the median daily duration was 3 min (IQR = 1–10 min) and the median number of distinct facilities visited was 1 (IQR = 1–1). Tap or borehole usage peaked between 17:00 and 19:00 across all districts (Supplementary Fig. 1b). Costs were consistent across districts: borehole usage cost 1,000 Ugandan shillings (UGX; equivalent to -US\$0.27) per household per month and taps cost 100 UGX (-US\$0.03) per jerrycan.

Overall, 24.3% (110/452) of participants visited both open-water sites and taps or boreholes, and 27.2% (123/452) visited neither. Usage

of open-water sites and taps or boreholes was weakly positively correlated (Spearman's rank correlation, $\rho_s = 0.14$; $P < 0.01$), as were usage durations ($\rho_s = 0.16$; $P < 0.01$). However, among participants with either open-water site or tap or borehole usage only (329/452), there was a negative correlation between the two durations ($\rho_s = -0.14$; $P < 0.01$).

Association of open-water contact with schistosome infection outcomes

At baseline, 47.6% (215/452) of participants were infected with *S. mansoni* based on Kato–Katz microscopy; after treatment with praziquantel, 17.0% (77/452) remained positive four to five weeks later, with 40.5% (183/452) reinfected at one-year follow-up. Median infection intensities among Kato–Katz-positive individuals were 96 eggs per gram (EPG; IQR = 36–264) at baseline, 36 EPG (IQR = 12–120) after treatment with praziquantel and 48 EPG (IQR = 24–126) at one-year follow-up. Full infection intensity data, including median EPG values at each time point, are reported in Supplementary Sections 1.1 and 1.2 and Supplementary Fig. 2. Duration of open-water contact was positively correlated with both baseline and one-year infection intensity ($\rho_s = 0.18$ ($P < 0.01$) and $\rho_s = 0.20$ ($P < 0.01$), respectively). Duration of tap or borehole usage was not associated with baseline infection intensity ($\rho_s = 0.03$; $P = 0.46$) and

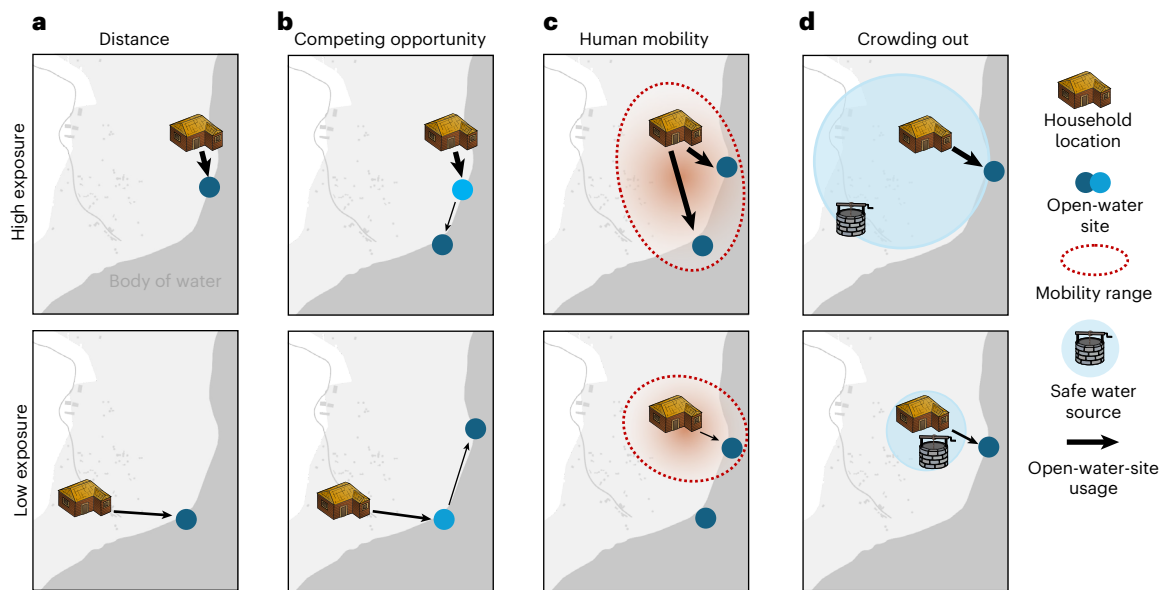


Fig. 3 | Schematic of different spatial factors influencing open-water-site usage. **a**, Open-water-site usage depends purely on household distance to the site. **b**, Open-water-site usage depends on household distance to the index site (dark blue) plus the distance to competing sites (light blue). **c**, Open-water-site usage depends on the range of human mobility. **d**, Open-water-site usage

depends on the proximity to safe water infrastructure, such as boreholes, whereby boreholes are able to reduce (that is, crowd out) open-water contact. Uganda waterbodies shapefile from World Resources Institute (<https://www.wri.org/>)⁷⁰ and base maps from OpenStreetMap (<https://www.openstreetmap.org>). House icon adapted from Firkin/Openclipart.

was only weakly positively correlated with one-year infection intensity ($\rho_s = 0.10$; $P = 0.04$).

Associations of age, gender and household distance with open-water contact

Participants ranged in age from 5–82 years and were 50% female (226/452). Age and gender had no significant effect on open-water site or tap or borehole usage (Supplementary Table 1). Occupation was relevant as fishermen and fishmongers were more likely to use open-water sites. Household access to an improved drinking water source was positively associated with both open-water and tap or borehole usage. Open-water contact peaked around age 25 years, whereas tap or borehole usage was largely stable across age (Supplementary Fig. 3). Within households where both a child and an adult participated, open-water contact and tap or borehole usage were each correlated between the two individuals ($\rho_s = 0.36$ ($P < 0.01$) and $\rho_s = 0.34$ ($P < 0.01$), respectively; $n = 191$).

Households in these fishing villages had a median distance of 219 m (IQR = 88–406 m) to the nearest open-water site and 335 m (IQR = 181–663 m) to the nearest tap or borehole. Fishermen lived significantly closer to open-water sites (median = 84 m; IQR = 58–190; $n = 41$) than other adults (median = 233 m; IQR = 74–400; $n = 186$; Wilcoxon rank-sum test statistic, $W = 2,599$; $P < 0.01$). Individuals travelled slightly further to visit open-water sites (median = 229 m; IQR = 109–416 m) than taps or boreholes (median = 172 m; IQR = 123–320 m), although the difference was not statistically significant ($W = 37,666$; $P = 0.15$). The nearest open-water site accounted for 47% (219/466) of all distinct sites visited and 55% of total contact duration (43,619/79,296 min), whereas the nearest tap or borehole accounted for 99.3% (149/150) of all distinct facilities used and 99.8% of total usage duration (16,417/16,447 min).

Association of human mobility with open-water contact

To capture the movement of individuals relative to their home location for visits to open-water sites or taps or boreholes, and for general purposes, we used the radius of gyration as our mobility metric (Methods). Aggregate mobility patterns were well approximated by a truncated power-law distribution ($R^2 = 0.73$ within a range of 1.3×10^1 – 1.1×10^4 m),

with a median radius of gyration of 356 m (IQR = 150–1,090 m; Fig. 2a). Although participants in Pakwach exhibited higher median movement than those in Mayuge or Buliisa, neither district-level nor occupational differences (for example, fishermen versus other adults) were statistically significant (Supplementary Section 1.3). We therefore grouped participants into district-specific mobility terciles for subsequent analyses (Fig. 2b). Tap or borehole visits were the most spatially focal (virtually all within 1 km), whereas open-water contacts extended to 10 km and general mobility to 100 km (Fig. 2c). Individuals living further from waterbodies (>1 km) appeared to have larger mobility radii, but this trend was not statistically significant (Fig. 2d and Supplementary Section 1.3).

Conceptual framework of the spatial factors influencing open-water contact

To identify key factors influencing open-water contact, we evaluated four spatial decay models of increasing complexity (Fig. 3). The baseline model hypothesized that open-water contact is solely determined by household distance to open-water sites (Fig. 3a). We sequentially expanded this framework to account for competing opportunities (where site distance rank influences choices between competing sites; Fig. 3b), individual mobility (Fig. 3c) and a crowding-out effect whereby tap or borehole usage reduces open-water contact (Fig. 3d). All spatial decay models were fitted as Bayesian nonlinear regression models predicting the probability of an individual using a specific open-water site or tap or borehole (equation (2) and Methods), with model selection performed using expected log pointwise predictive density (ELPD) differences.

Spatial decay models of open-water usage

An exponential spatial decay model (Supplementary Fig. 4), corresponding to Fig. 3a, best approximated open-water-site usage, outperforming a simple logistic regression using household distance (ELPD difference = 23.7; $P < 0.01$), particularly at small spatial scales (<100 m; Supplementary Fig. 5). Distance-based models also outperformed logistic regression models using sociodemographic and behavioural proxy variables (Supplementary Table 2).

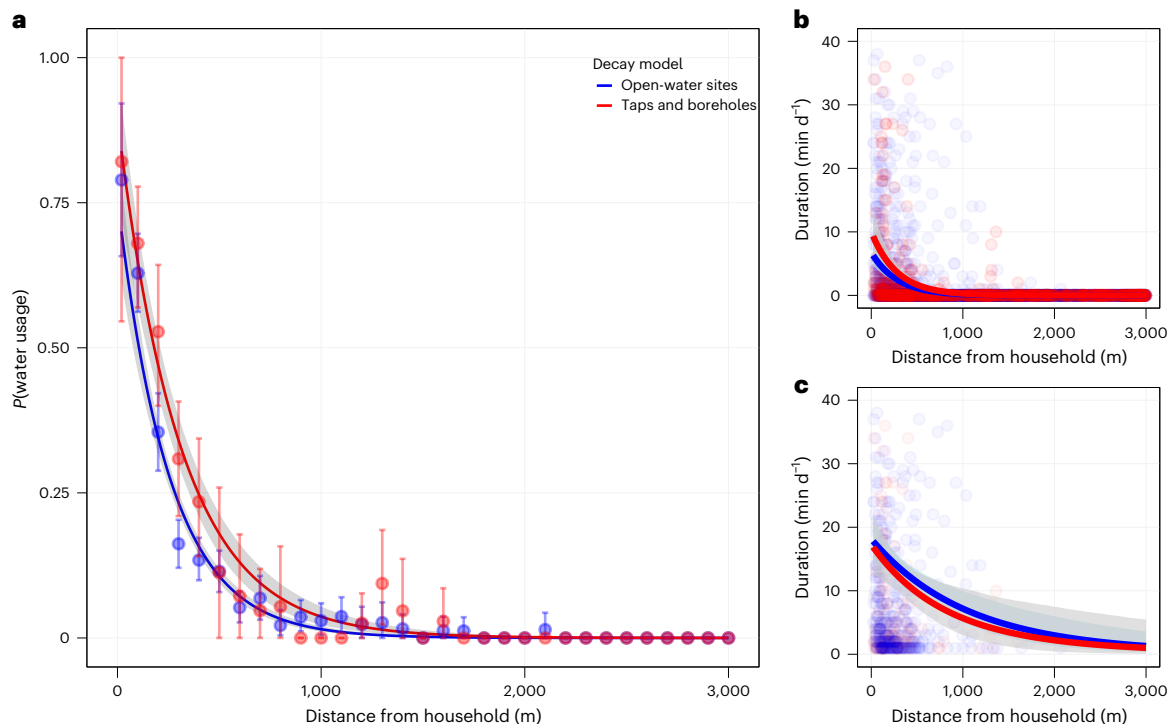


Fig. 4 | Spatial decay baseline models. **a**, Individual-level spatial decay models (referred to as global models) predicting the probability of open-water site and tap or borehole usage over household distance. The dots represent the data, with error bars indicating 95% CIs obtained via non-parametric bootstraps with replacement. The error bars represent the 2.5th and 97.5th percentile ranges, calculated using 2,000 replicates and plotted in 100-m increments. The points

represent the respective means. **b,c**, Spatial decay models predicting the duration of open-water-site contact and tap or borehole usage over household distance among all individuals (**b**) and restricted to individuals using either open-water sites or taps and boreholes, respectively (**c**). The solid lines and shaded grey areas indicate posterior means and 95% CrIs, respectively. In **b** and **c**, the points represent the raw data.

The global spatial decay model of open-water-site usage (Fig. 4a) had a stratified fivefold cross-validated area under the receiver operating characteristic curve (AUROC) of 0.86 (95% confidence interval (CI) = 0.85–0.88). For individuals living close to open-water sites (20 m), the predicted probability of usage was 70% (95% credible interval (CrI): 61–78%), decreasing to 51% (95% CrI = 46–56%) at 100 m, 11% (95% CrI = 9–12%) at 500 m and 1% (95% CrI = 0.7–1.4%) at 1,106 m (Fig. 4a).

Tap or borehole usage exhibited a broadly similar spatial decay, although the decrease was somewhat steeper (Fig. 4a), with a higher AUROC of 0.92 (95% CI = 0.90–0.94). At 20 m, the predicted probability of usage was 84% (95% CrI = 72–93%), 14 percentage points higher than for open-water sites, decreasing to 65% (95% CrI = 57–71%) at 100 m, 18% (95% CrI = 15–21%) at 500 m and 1% (95% CrI = 0.5–1.7%) at 1,413 m, compared with 1,106 m for open-water-site contact (Fig. 4a).

We additionally used exponential spatial decay models to predict the duration of open-water contact and tap or borehole usage (Fig. 4b,c), with moderate predictive performance ($\rho_s = 0.36$ for open-water contact and $\rho_s = 0.47$ for tap or borehole usage) and similar decay rates for the duration and probability of any usage (both decay rate parameters $b_1 = 0.004$). Given the substantial number of non-users, we refit the duration model restricted to participants with any use ($n = 275/452$ for open-water sites; $n = 135/452$ for taps or boreholes). For these individuals, decays were less pronounced and predicted durations remained positive, even at 2,000 m (Fig. 4c).

Spatial decay models of open-water contact with additional covariates

We tested whether including additional covariates improved the ability of spatial decay models to predict open-water contact and tap or borehole usage. District-specific distance decays performed

best, outperforming the respective global models (ELPD differences of 13.0 ($P = 0.04$) and 10.1 ($P = 0.09$); Supplementary Table 3), with AUROCs of 0.87 (95% CI = 0.85–0.89) and 0.92 (95% CI = 0.90–0.94) for open-water contact and tap or borehole usage, respectively (Supplementary Fig. 6). The performance of the district-specific models was similar to that of models including gender, activity match, age or safe water type as covariates (all ELPD difference P values > 0.05 ; Supplementary Table 3). We used the district-specific model (Supplementary Fig. 7) as our benchmark for all subsequently introduced models. The ability of this model to predict population-wide open-water contact patterns and site-level traffic using building footprints as a proxy for population density is demonstrated in Supplementary Section 1.9 and Supplementary Fig. 8.

Spatial decay models of open-water contact incorporating competing opportunities, human mobility and crowding out

To further refine the spatial decay model, we evaluated whether incorporating site competition, human mobility and crowding out (hypotheses shown in Fig. 3b–d) improved performance over the distance-based benchmark. The competing opportunities model used relative distance ranks alongside absolute distance, so that comparatively closer sites reduced the usage of more distant sites. This significantly improved the predictive performance over the benchmark (ELPD difference = 52.1; $P < 0.01$; AUROC = 0.87; 95% CI = 0.86–0.89), with the monotonic decay in Fig. 5a confirming that the usage probability decreased as the number of closer competing sites increased, even after accounting for absolute household distance. A simpler rank-only model without absolute distance performed substantially worse (ELPD difference = –111.6; $P < 0.01$; Supplementary Fig. 9).

We also examined whether higher mobility was associated with visiting more distant open-water sites (Fig. 3c). Individuals in the

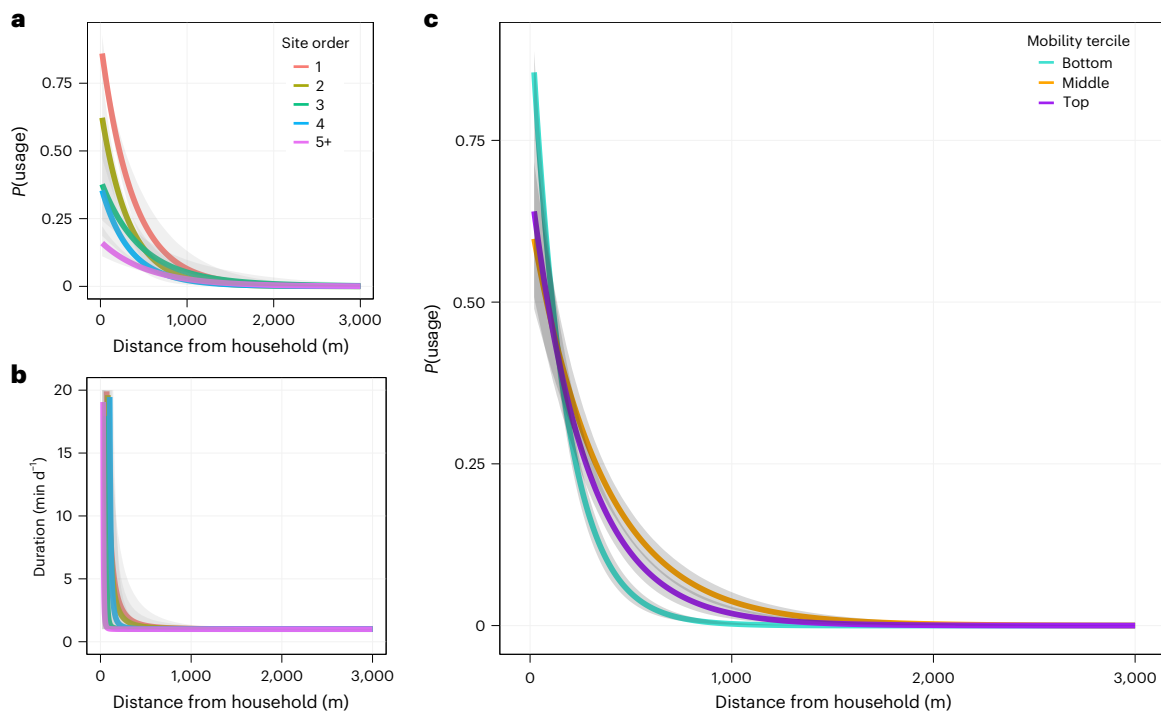


Fig. 5 | Spatial decay models of competing opportunities and human mobility. **a**, Spatial decay model of open-water-site usage based on relative distance to competing open-water sites. For each individual, all open-water sites were ranked by increasing distance from the household, where 1 is the nearest and 5 is any site farther than the fourth-nearest. A separate distance-decay function was then

fitted for each rank category. **b**, Spatial decay model predicting open-water-site duration based on site order. **c**, Spatial decay model predicting open-water-site usage probability based on district-specific mobility tertile. The solid lines indicate posterior means and the shaded grey areas represent 95% CrIs.

middle mobility tertile maintained the highest open-water contact probabilities at greater distances, whereas those in the high and low tertiles showed steeper decays (Fig. 5c). The mobility-adjusted model did not significantly outperform the benchmark for either open-water contact (ELPD difference = 14.2; $P = 0.20$) or taps or boreholes (ELPD difference = -11.6; $P = 0.07$).

The crowding-out model failed to outperform the benchmark for either site usage probability (ELPD difference = -2.4; $P = 0.22$) or duration (ELPD difference = -0.02; $P = 0.99$). Even in a counterfactual scenario assuming this unsupported model held true, we estimated that <2% of individuals (95% CrI = 0–5%) would cease open-water contact due to tap or borehole usage, with a maximum population-level reduction in open-water contact duration of 23% (95% CrI = 19–37%) under universal adoption, and no significant variation in crowding out by contact type (occupational, domestic or recreational; Supplementary Section 1.6 and Supplementary Figs. 10 and 11).

Model robustness checks

We assessed the minimal data needs and robustness of our main (district-specific open-water contact) model to GPS recording frequency and buffer size. As few as 15 individuals recorded over ten days were sufficient to estimate the spatial decay (Supplementary Fig. 12a), and the decay parameters b_0 and b_1 were robust to removing contact events of <5 or <10 min duration, respectively (Supplementary Section 1.7 and Supplementary Fig. 13). Using buffer sizes of 20 and 30 m yielded similar results (Supplementary Fig. 14). Males appeared somewhat less compliant in wearing the GPS loggers than females ($P = 0.01$; Supplementary Section 1.8 and Supplementary Table 5) and the district-specific model was modestly less accurate for fishermen and adults ($P < 0.01$ and $P = 0.04$, respectively; Supplementary Section 1.10 and Supplementary Table 6). Despite this, the global model, although not the best-performing model, showed high transportability across districts when each was held out in turn (AUROC = >0.82; Supplementary Fig. 12c).

Using model-based open-water site contact probabilities to inform open-water contact in IBMs

We incorporated a spatial decay function of open-water contact into an IBM (Methods), having confirmed that model-based estimates of open-water contact were significantly correlated with baseline infection intensity ($\rho_s = 0.18$; 95% CI = 0.09–0.26) and one-year infection intensity ($\rho_s = 0.20$; 95% CI = 0.11–0.27; for validation of GPS-derived against self-reported open-water contact; Supplementary Section 1.4).

Using district-specific decay model parameters, we investigated four visitation rules governing individual contact with open-water sites within the IBM to understand how the choice of visitation rule influenced predictions of infection one year after treatment (Fig. 6). The rules, spanning a range of assumptions about how individuals distribute contacts across sites, were: (1) visit the nearest site with 100% probability; (2) visit the nearest site with probability P from the spatial decay model; (3) apply Bernoulli trials with probability P across all sites and visit one success at random; and (4) follow the same rule as (3) but visit all sites assigned a success.

The IBM yielded the total number of visits for each site over a one-year simulation period (Extended Data Fig. 1). There was between-district heterogeneity in the percentage of used open-water sites with a presence of *Biomphalaria* intermediate hosts ranging from 63.6% (Buliisa) to 100% (Pakwach), with no obvious relationship between the number of contacts and snail presence. Rules (3) and (4) predicted more sites being used than rules (1) and (2), which restricted visitation to the nearest site only. More than half of all mapped sites were predicted to be unused under all rules (Supplementary Section 1.5).

We quantified the human-to-snail transmission input by estimating cumulative miracidial contamination introduced through human water contact and expressed this as a percentage of the most contaminated site per district (Supplementary Fig. 15). Contamination distributions broadly mirrored visit distributions, although site rankings

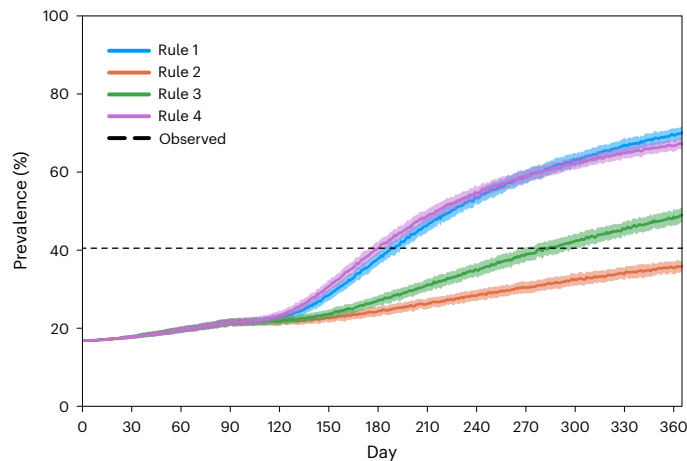


Fig. 6 | Predicted infection prevalence based on four different open-water-site visitation rules. The approximate start of the simulations is the post-treatment infection prevalence (16.9%). The dashed black line indicates the reinfection prevalence (40.5%) observed one year after baseline. The shaded areas represent 95% CIs, calculated based on 100 replicates for each decision rule for 365 days. The main text provides detailed information on each rule.

occasionally differed between the two metrics, indicating that high visit counts do not always correspond to the highest transmission input.

The IBM predicted a one-year post-treatment infection prevalence of 70.1% for rule (1), 35.8% for rule (2), 49.0% for rule (3) and 67.2% for rule (4), with rules (2) and (3) closest to the observed prevalence of 40.5% (183/452). Sensitivity analyses on the aggregation parameter (Supplementary Fig. 16) showed that the ordering of rules was robust, although differences between rules narrowed at lower parameter values, which are more appropriate for low-prevalence settings than ours.

Discussion

The WHO schistosomiasis guideline emphasizes the need for spatially focal interventions to reduce schistosome exposure in endemic areas¹³. We studied open-water site contact and tap or borehole usage patterns of 452 children and adults in rural Uganda over a ten-day period, identifying sites with the highest probability of usage and frequency of contact. Spatial decay models of open-water contact, constructed from wearable GPS logger data, showed high predictive performance and informed more realistic assumptions in IBMs of schistosome transmission.

Individual-level spatial decay models informed by household distance and district alone predicted open-water contact and tap or borehole usage well (AUROCs of 0.87 and 0.92, respectively), and a simpler purely distance-based model still had high transportability across diverse Ugandan districts (AUROC > 0.82). This contrasts with previous studies using aggregate cell-phone mobility data, which found that gravity models (spatial decay models similar to ours) performed poorly in rural settings^{29,30}. Previous work predicting open-water contact used coarse mobile phone data^{10,12}, no mobility data^{11,19} or small GPS logger samples^{27,31}. Our models demonstrate that simple exponential spatial decay functions can accurately approximate fine-scale water contact patterns relevant to schistosomiasis. A similar decay has been proposed for water usage in trachoma³² but not rigorously evaluated, highlighting the need for future research on whether spatial decay models generalize to other environmentally mediated infectious diseases.

Integrating spatial decay into schistosome transmission models reflected the biological realism of complex human water contact². Previous theoretical work has shown that schistosome transmission depends not only on the total number of open-water contacts, but also on their distribution across sites³³. We demonstrated that this distribution can be approximated using simple household-to-site assignment

rules learnt from wearable GPS loggers. Rules informed by spatial decay probabilities and restricted to a single site visit per day produced one-year post-treatment infection prevalence estimates closest to the observed prevalence, outperforming the most common spatial assumption of assigning all contacts to the nearest open-water site^{8,16}. Incorporating space explicitly therefore captures the heterogeneity and focality of human exposure, providing a more reliable foundation for understanding district-level infection patterns.

Our transmission model provides the information needed for a localized decision framework targeting focal interventions at both humans and snails. Combining site-specific visit counts with data on snail presence and miracidial contamination allows for directly inferring transmission potential at each site and identifying priority sites for intervention depending on the degree of heterogeneity. Highly skewed contamination distributions support focal mollusciciding³⁴, other biocontrol methods³⁵ or intensified snail surveillance, whereas more uniform distributions suggest the need for broader coverage. A mismatch between visit distributions and miracidial contamination at the site level can identify high-burden individuals for test-and-treat strategies³⁶ and facilitate a comparison of human and snail forces of infection to determine whether the same sites drive both sides of the transmission cycle.

The special decay model results suggest that safe water provision alone is unlikely to eliminate open-water contact. Crowding-out models did not outperform purely distance-based models, and even under counterfactual scenarios assuming crowding out existed, fewer than 2% of participants would completely divert their open-water contact to taps or boreholes. This aligns with Knopp et al.¹⁵ and the WHO guideline¹³, which anticipate only modest WASH effects, rather than the larger reductions implied by observational meta-analyses¹⁴. Further experimental work on WASH delivery models (household versus public provision)³⁷, infrastructure types (boreholes, standpipes, water kiosks and so on) and cost structures is needed to identify conditions under which WASH might meaningfully divert high-risk water contact activities, even if total open-water contact is not substantially reduced.

Despite interest in identifying mobile individuals who may be missed by MDA programmes^{38,39}, we found few clear sociodemographic correlates of human mobility. Adults had larger movement ranges than children, although median mobility did not differ significantly, and fishermen were not found to be more mobile than other adults, in contrast with previous suggestions⁴⁰. Spatial decay models incorporating mobility did not outperform simpler distance-based models, indicating that mobility patterns cannot serve as a direct proxy for open-water contact^{10,11}.

Spatial decay models have practical applications across multiple elements of schistosomiasis control¹³. As essentially all open-water contact and tap or borehole usage occurred within 1 km of the homes of individuals, exposure reduction should focus locally, limited to home and neighbouring villages. Our models can predict the number of people using each open-water site, guiding focal mollusciciding of the most visited sites⁴¹. Notably, radius control should extend to 1 km from households, larger than the 500 m suggested by the WHO⁴¹, to cover all locally relevant sites. GPS logger data from as few as 15 individuals over ten days were sufficient to estimate spatial decay, demonstrating logistical feasibility. We also showed that publicly available building footprints or WorldPop data^{42,43} can approximate population-level open-water contact, although generalizability will depend on local infrastructure, social norms and the representativeness of the GPS sample and observation season.

The strengths of our study include: the large-scale collection of wearable GPS logger data; extensive model investigation, including transportability across districts; proposed theories of change for spatial decay models; and validation with reinfection data and spatially explicit IBMs. Our study has several limitations. We focused on the spatial dynamics of open-water contact and did not model

temporal patterns beyond descriptive reporting of diurnal usage. The ten-day observation window may not be representative of general usage patterns and models may not generalize to the rainy season when open-water site distributions and accessibility differ²⁶. However, the dry season focus captures a key transmission period when contact is highest^{2,16}. As open-water contact is inferred from GPS logger positions and not directly observable, there is a risk that our estimates of open-water contact have noise (although we have no reason to believe they are systematically biased). From the qualitative data we collected, we did not receive any reports (although these were not elicited in a structured manner) that individuals changed their routine behaviours while wearing the loggers. We removed the first day of observation to guard against Hawthorne effects and found that decay parameters were robust to the exclusion of short-duration contact events (<5 min). Our 20 m buffer was stringent relative to previous studies^{27,31}, yielding conservative contact estimates. Nonetheless, the observed median daily contact duration of 3 min is consistent with an observational study of 120,000 open-water contacts in Senegal (4.3 min)²² and another study from Kenya showing that the average number of open-water sites typically visited by people was less than one²³. As our study mapped open-water sites up to 26 km away, and with >99.5% of all open-water contact occurring within 3,000 m, unmapped distant sites are unlikely to have affected the findings, although some local contacts may have been missed if sites were not captured during mapping. All study communities were within 2 km of waterbodies and our findings may thus not apply to communities further from the shoreline. Spatial decay models assign identical predictions to all household members, limiting the investigation of intra-household dynamics, but incorporating sociodemographic covariates did not improve model performance over a purely spatial model. Due to the nature of the GPS logger data, we had no information on specific activities; activity-level data are better captured through surveys or direct observations^{2,16}. Validation of GPS-derived against self-reported open-water contact yielded a weak positive correlation, partly reflecting imperfect alignment between the GPS observation period and survey recall window. Future studies should use directly comparable measurement periods.

In conclusion, we provide individual-level spatial decay models of human water contact relevant for modelling schistosome transmission. Future research should assess whether similar models generalize to other waterborne diseases, recognizing pathogen-specific exposure pathways, and should collect GPS logger data across diverse geographical settings to establish the broader generalizability of observed spatial decays. With further evaluation, our models could serve as planning tools for targeting focal interventions, including WASH, mollusciciding and MDA prioritization.

Methods

Ethical approval

Data collection and use were reviewed and approved by the Oxford Tropical Research Ethics Committee (509-21), Vector Control Division Research Ethics Committee of the Uganda Ministry of Health (VCDREC146) and Uganda National Council for Science and Technology (HS 1664ES). Written informed consent was obtained for all study participants, with adults consenting on behalf of children and children providing verbal assent.

Study context

In sub-Saharan Africa—the region with the lowest access to safe drinking water—408 million people rely on unsafe sources such as rivers, lakes and ponds⁴⁴. This study focuses on Uganda, where <10% of people in rural areas are estimated to have access to safely managed drinking water⁴⁴ and where the national prevalence of schistosomiasis has been estimated to be 26% using point-of-care circulating cathodic antigen tests⁴⁸. Despite over 13 rounds of MDA since 2003, repeated MDA in Uganda has failed to achieve the WHO targets of morbidity control

and elimination as a public health problem⁴⁵. The 2022 WHO schistosomiasis guideline highlights the need for multifaceted control, such as combining MDA with behaviour change and WASH, to achieve the WHO 2030 control targets¹³. SchistoTrack collected comprehensive environmental, behavioural and spatial data to inform more integrated, focal schistosomiasis control programmes. The study focused on moderate- and high-endemicity villages in western and eastern Uganda, where the prevalence of schistosomiasis, as measured by stool microscopy, was 43%². The study took place across three Ugandan districts, Pakwach, Buliisa and Mayuge, which are located along the River Nile, Lake Albert and Lake Victoria, respectively²⁵. These areas were selected to represent diverse climates (tropical rainforest and tropical savannah climates), waterbody types (lake and river settings) and tribal and religious groups. All villages were <2 km away from waterbodies, and people engaged primarily in subsistence farming or agriculture, with 47% reporting open-water contact in the past week².

Study design and participant sampling

The GPS logger study was nested within the SchistoTrack cohort study baseline, which took place between January and February 2022 and enrolled 1,459 households across 38 villages ($n = 2,885$), approximating 40 households per village². Participants were randomly selected from village registers or MDA records (see Puthur et al.²⁵ for details). Of those 38 villages, 12 (four in each district) were selected for the GPS logger study based on their levels of open-water contact and occupational fishing, determined predominantly by the presence of a beach or landing site. None of the selected villages had piped water (that is, individuals had no access to safe drinking water within their households). A maximum of 50 participants from 25 households per village were selected among SchistoTrack participants, with a target of one adult–child pair (with adults defined as people ≥ 18 years of age and children defined as those 5–17 years of age) per household²⁵. This study excluded children aged <5 years, as there was no approved paediatric praziquantel formulation for this age group at the time of the study. For this nested cross-sectional study, SchistoTrack participants were sampled to ensure a balance of age and gender, and fishermen were oversampled to ensure a diverse representation of open-water contact and mobility patterns. We excluded participants who were physically immobile, as defined by the participant, and thus unable to have open-water contact. We also excluded children who were sent away to boarding school and were unavailable during the GPS logger study period. All participants were unaware of their schistosome infection status when they entered the GPS logger study.

Household survey data

Before the GPS logger study, questionnaires to collect sociodemographics, biomedical variables, WASH and environmental variables and self-reported open-water contact patterns were administered to the household head of all 1,459 households that were part of the larger SchistoTrack study. Questionnaires were translated into local languages and administered digitally by two surveyors (one recruited centrally from Kampala and one recruited from within the district) trained by G.F.C. and M.T.E. The data were collected via a tablet device (Lenovo TB-8505F) and the Open Data Kit platform (using Open Data Kit Collect application version 2022.4). To ensure the quality of the survey data, questionnaires were piloted using field tests and mock interviews; response constraints were implemented to avoid implausible response values as far as possible; and standardized interview protocols were used to ensure consistent interview procedures were followed by surveyors. At the end of the household interview, one adult and one child per household were selected for clinical assessments by the household head. From these clinical participants, a subset was subsequently selected for inclusion in the GPS logger study.

Community sensitization, consent and remuneration

G.F.C., M.T.E. and F.R. conducted community workshops in each village to explain the study objectives and address any questions or concerns. As part of the standard written informed consent for recruitment into SchistoTrack, consent was obtained from adult participants and parents or guardians for children under 18 years in the local language, specifically for the GPS logger study. Children provided verbal assent and, where possible, written assent was also collected. Participants received a nominal remuneration equivalent to about one day of wages (10,000 UGX; -US\$2.73) in addition to any incentive in the study upon returning their GPS loggers at the end of the observation period.

GPS logger settings and procedures

Wearable GPS loggers (i-gotU GT-600; Mobile Action Technology, Taiwan) were used to record the location data of participants at two-minute intervals between 05:00 and 20:00 local time over approximately ten consecutive days. This time window was chosen to capture activities during the day. Loggers were turned off from 20:00–05:00 to save battery, using scheduled control, as loggers were not recharged in the field by participants or the study team. Also to save battery, all loggers were set to motion activation such that loggers turned off when participants did not move. Button controls on devices were disabled, so participants were unable to switch them off. Devices were placed in waterproof pouches and worn around the neck using a lanyard (Supplementary Fig. 17). Children and adults in the same household always received different-coloured lanyards to minimize the risk of accidentally mixing up the loggers. Adults were asked to help children with the day-to-day wear of the loggers. Participants were instructed to wear the devices during all daily activities, including during open-water contact, and to remove them only at night.

Participant flow

A participant flowchart is shown in Supplementary Fig. 18. A total of 215 loggers were purchased for the study, with the aim of distributing 200 loggers across three districts sequentially (that is, 600 in total). The total number of recruited participants was 585—slightly below the target of 600. This was mainly due to device failure and unreturned loggers in Pakwach. After excluding participants with fewer than two complete days of GPS data—excluding the first day of recording for all participants to reduce the Hawthorne effect⁴⁶—we included a final analytic sample of 452 participants in the study.

GPS logger testing and validation

We tested the accuracy of GPS loggers *in situ*. During fieldwork, ten loggers were placed next to each other on the ground for 10 min in an unobstructed area in Buliisa to estimate GPS position error in the field. Based on this experiment, we estimated that GPS loggers were accurate within 8.6 m. GPS precision was estimated using the `uere` function in the `ctmm` package in R⁴⁷. Our estimate of 8.6 m aligned with a previous study, which estimated the accuracy of i-gotU GT-120 under different levels of obstruction and found that the mean location error was <10 m and that substantial cover moderately increased the location error by 2.2 m⁴⁸.

GPS logger data retrieval

For 94 participants, no GPS data could be retrieved as loggers suffered water damage despite the use of protective pouches and the manufacturer's indication that devices were fully waterproof. Multiple retrieval strategies were employed: metal contacts were cleaned with ethanol in the field; loggers were transported to Oxford for a second cleaning attempt; and 36 devices with persistent data loss were opened at the University of Oxford Materials Science Department, where a research technician resoldered the GPS chips onto functioning logger boards. This recovered data from 24 of the 36 affected loggers.

Self-reported wear compliance

Wear comfort was assessed when participants returned GPS loggers. As not all devices were returned immediately following the end of the study period by the participants themselves, these data were available for 87% of participants (395/452). Among those participants, 83% (328/396) reported that wearing the logger was comfortable. Other participants self-reported that loggers caused perceived chest pain (34), body weakness (5), nausea (1) or headaches (1), which were addressed through sensitization.

Mapping of village infrastructure, open-water sites and taps and boreholes

As part of the study, F.R., M.T.E. and (initially) G.F.C. comprehensively mapped the locations of all public taps and boreholes used by the study communities that were relevant for schistosome transmission, assisted by the village chairman or a village health team member. As part of this mapping, the type (tap or borehole), GPS location and state of each facility were assessed. Trained malacologists from the Division of Vector-Borne and Neglected Tropical Diseases at the Uganda Ministry of Health mapped all open-water sites used by study communities, including sites outside village perimeters but accessed by community members with the help of local guides, such as village chairmen or community health workers. This comprehensive mapping approach ensured the inclusion of all relevant open-water sites used by the study population, regardless of their proximity to residential areas, as is described in detail by Iacovidou et al.²⁶.

Deriving open-water contact and tap or borehole usage patterns

A contact event with an open-water site and tap or borehole was defined as being within a 20 m buffer of the GPS location of an open-water site or tap or borehole. We chose 20 m to account for possible GPS position errors while minimizing overestimation of open-water contact. This buffer was conservative compared with previous GPS logger studies; Eyre et al.²⁷ used a 30 m buffer around the shoreline, whereas Seto et al.³¹ used a 100 m buffer around the shoreline. We used a larger buffer of 30 m as a robustness check, with the results reported in Supplementary Fig. 14. GPS locations of sites and taps or boreholes were recorded via tablet (Lenovo Tab M8 (3rd Gen); Lenovo Group; China) with 5–10 m manufacturer-reported accuracy, and GPS logger points were reported with 8.6 m accuracy, as described above. All contact events with open-water sites and taps or boreholes were identified using recursion analysis—a well-established method in movement ecology^{49,50}—and implemented in the `recurse` package in R⁵¹. The output from the recursion analysis was a list of all distinct open-water contact and tap or borehole usage events with an associated timestamp, number of revisits, contact duration and open-water site and tap or borehole identifier (Fig. 1e). Compared with simply counting the number of points within each buffer area, recursion analysis has the advantage of reconstructing an approximate linear trajectory. It can therefore identify visits with durations shorter than the GPS logger sampling frequency. This is particularly important given that water contacts are typically short²².

As the SchistoTrack protocol did not allow sites to be larger than 15 m, even when they were directly adjacent, we grouped such nearby locations into a single open-water site (or a single tap or borehole for adjacent taps and boreholes) for the analysis, following the methods used by Iacovidou et al.²⁶. The density-based spatial clustering of applications with noise algorithm was used and the optimal number of clusters was determined based on the gap statistic, separately by district⁵². This collapsed the 143 open-water sites and 63 taps and boreholes into 69 and 32 clusters, respectively, to which we then assigned the mean GPS cluster coordinates.

Sanitation infrastructure

This analysis did not focus on the use of public latrines, even though it would have been possible to derive this information from the GPS data

using the same methods as for open-water site and tap or borehole usage. There were two reasons for this. First, field visits to all public latrines by M.T.E. and F.R., and initially by G.F.C., revealed that most were locked (usually a community member held the keys) and not regularly used by the community. Second, among the 452 GPS logger participants, 343 (75.9%) lived in households that had a private latrine (that is, a flush or pour-flush toilet, a covered pit latrine with or without privacy or a composting toilet). Thus, most households could rely on private latrines, which provide more privacy and convenience compared with public latrines^{53,54}. However, private latrine usage events could not be derived from GPS logger data due to the unknown precise location of these latrines.

Dataset and variables

Dataset. The dataset used to estimate spatial decay models was generated by calculating the Euclidean household distance to all open-water sites and taps or boreholes (both used and unused) within the same district (Fig. 1e). This covered distances of up to 11, 17 and 26 km for Pakwach, Buliisa and Mayuge, respectively.

Outcomes. For the models estimating any usage of open-water sites or taps or boreholes, the outcome was a binary indicator of whether an individual used an open-water site or a tap or borehole. For the model estimating duration of usage, the outcome was a continuous variable indicating the number of minutes per day each individual had contact with each mapped open-water site or tap or borehole. Duration, in minutes per day per individual, was calculated by dividing the total duration of contact with each open-water site or tap or borehole by the number of distinct calendar days with GPS data for each individual.

Covariates. The main predictor variable was household distance to each open-water site or tap or borehole. We selected additional predictor variables based on their relevance for open-water contact. The following individual-level variables were used: age, gender, occupation, self-reported open-water contact, water contact activity match and mobility tercile. Age was coded as a binary variable (<15 or ≥15 years). This cut-off was chosen based on our previous analysis of the SchistoTrack data, which demonstrated that open-water contact levels were low in individuals aged <15 years and increased substantially thereafter, suggesting that 15 years of age was a natural cutoff². We also used gender (male or female), reported by the household head, as a covariate. Individual-level occupation was also used as a covariate because occupation is an important determinant of open-water contact². The household head (aged ≥18 years) reported the occupation of all household members. The occupation categories were fishing, fishmongering, farming and other. A self-reported open-water contact variable, described in detail in Reitzug et al.², was also included. A binary open-water contact activity match variable was generated for each individual–site pairing. Whenever an individual conducted any of the 11 domestic, recreational or occupational open-water contact activities, and whenever the same activity was reported to be performed at a specific site, based on reports by the local guide, this was defined as an activity match. The 11 open-water contact activities included: collecting drinking water; washing clothes with soap; washing clothes without soap; bathing with soap; bathing without soap; washing jerry cans or household items; collecting papyrus; fishing; fishmongering; collecting shells; and swimming or playing. All individuals who reported collecting drinking water were also assigned to having an activity match with taps or boreholes, as this activity was amenable to being conducted at a tap or borehole. Apart from the activity match, open-water contact activities of an individual were not considered in this analysis, as the GPS data did not contain any definitive information on the specific open-water contact activity an individual engaged in. We also used an individual's mobility tercile as a covariate (see below). At the household level, aside from distance

to the closest sites and taps or boreholes, a self-reported variable for whether the household used a safe drinking water source was also used. Taps and boreholes were counted as safe, whereas open freshwater sources such as swamps or lakes were counted as unsafe. To account for geographic, cultural, behavioural and environmental factors that differed across study locations, a district-level categorical variable was included in some models. Households paid different prices for tap usage compared with boreholes (see results); therefore, we also generated a variable indicating whether each public water supply was a tap or a borehole.

Schistosome reinfection. To test associations between water contact and (re)infection intensity, we relied on schistosome infection data from the SchistoTrack study. All participants were tested for schistosome infection using Kato–Katz microscopy at baseline, four to five weeks later following treatment with praziquantel and at one-year follow-up. Baseline testing took place before the GPS logger hand-out and the study team was unaware of the Kato–Katz results. Treatment with praziquantel occurred irrespective of infection status and participants remained unaware of their infection status throughout the study. Infection intensity was measured in EPG of stool. The one-year post-treatment infection measurement was chosen as this is the interval for MDA in endemic areas and it is frequently used as a time frame to assess reinfection^{13,55,56}. Among participants, 0.4% (2/452) were missing Kato–Katz results at baseline, 3.1% (14/452) were missing them at treatment follow-up and 20% (91/452) were missing them at one-year follow-up. For those individuals, we imputed the reinfection intensity based on the baseline infection intensity using the multivariate imputations by chained equations technique, implemented in the *mice* package in R⁵⁷.

Human mobility measures

We aimed to compare human mobility patterns across different activity types: visits to open-water sites or taps or boreholes and overall movement. To do so, we selected the radius of gyration $R_{g,i}$ as our metric, as it measures the Euclidean typical displacement of individual i from their household location and is widely used in human mobility modelling^{58–60}. This choice ensured consistency with our use of Euclidean distances for calculating household distances to visited open-water sites and taps or boreholes. $R_{g,i}$ was computed using the following formula:

$$R_{g,i} = \sqrt{\frac{1}{N} \sum_{l=1}^N (\vec{r}_l - \vec{r}_h)^2} \quad (1)$$

where $(\vec{r}_l - \vec{r}_h)^2$ represents the squared Euclidean distance from location l to the household location h .

Spatial decay models

We modelled open-water site and tap or borehole usage as a function of Euclidean household distance, motivated by evidence that open-water contact and schistosome infection are waterbody distance dependent^{2,17,23}. Our previous work using self-reported open-water contact and closest site assignment suggested that the water contact over household distance was approximately linear². The modelling framework proposed here differs from previous studies in that it aims to estimate the usage of each open-water site and tap or borehole, as opposed to just the closest site. Our spatial decay models take the general form:

$$P(\text{response}_{i,k} = 1) = b_0 \times f(d_{i,k}) \quad (2)$$

where P is the probability that individual i uses a specific open-water site and tap or borehole k . Here $f(d_{i,k})$ represents the spatial decay function with d being the household distance to k . The binary outcome

(response_{*i,k*}) is modelled using a Bernoulli distribution with an identity link function to represent probabilities directly. All spatial decay models were fitted as Bayesian nonlinear regression models.

The sequential model-building process we used was as follows. First, we determined a suitable spatial decay function $f(d)$ in equation (2). We tested whether an exponential decay or a power-law decay—two decay functions used commonly in human mobility modelling⁵⁸—or a Hill function (which includes more parameters to control the curvature) best fit the data (Supplementary Table 7). As models were fitted as Bayesian nonlinear models, the best decay function was determined via ELPD—a Bayesian leave-one-out measure of out-of-sample predictive fit⁶¹ (see below for more details). Second, we tested whether the inclusion of covariates, added one variable (j in equation (3)) at a time (from district, gender, age (<15 or ≥15 years), occupation category or activity match), improved the model fit compared with the base model without additional covariates (global model; equation (2)).

$$P(\text{response}_{i,j,k} = 1) = b_{0j} \times f_j(d_{i,j,k}). \quad (3)$$

We then extended equations (2) and (3) to evaluate the influence of spatial distance, human mobility and usage of safe water infrastructure on human contact with open-water sites, as shown in Fig. 3. For competing opportunities (Fig. 3b), we drew up Stouffer's intervening opportunity framework⁶² to develop a spatially explicit model estimating the decay separately for the first, second, third, ..., n th closest site. As <1% (4/452) of individuals visited more than five distinct open-water sites, all sites beyond the fifth nearest were collapsed into a single category. For human mobility, the effects (Fig. 3c) were assessed by estimating decay separately per mobility tercile, represented by j in equation (3). Crowding out (Fig. 3d) was assessed using equation (5).

Analogously to the model predicting the probability of open-water-site usage in equation (2), we used the same modelling approach to predict tap or borehole usage, with $P(\text{response} = 1)$ representing the probability of using a tap or borehole. We also used spatial decay models to predict the duration of open-water site and tap or borehole usage. These models took the following form:

$$E(t_{i,k}) = b_0 \times f(d_{i,k}). \quad (4)$$

We modelled equation (4) using a zero-inflated negative binomial model:

$$t_{i,k} \sim \begin{cases} 0, & \text{with probability } \pi_{i,k}, \\ \text{NegBin}(\mu_{i,k}, \phi), & \text{with probability } 1 - \pi_{i,k}, \end{cases}$$

where $\mu_{i,k} = b_0 \times f(d_{i,k})$ is the mean duration for individual i at site k , and ϕ is the dispersion parameter, which allows the negative binomial to model strong overdispersion and a heavy right tail in visit durations. The zero-inflation probability $\pi_{i,k}$ accounts for excess zeros beyond those produced by the negative binomial component. Due to the disproportionate influence that a very small number of extreme duration values can have on the dispersion parameter ϕ , and to obtain parameter estimates that were more representative of typical open-water contact behaviour, we capped all durations above the 99.9th percentile at that percentile value (>139 min d⁻¹ set to 139). This transformation affected only 0.1% of observations and preserved the rank order of almost all data while limiting the undue influence of extreme values on the zero-inflated negative binomial fit. For exponential duration decay models, we used an uninformative prior, Unif(0, 1), for b_1 and a normal prior, Normal(1, 9), truncated at zero and 30, for b_0 . This prior was chosen to match the mean and standard deviation of the data.

We also built spatial decay models in which the usage of taps or boreholes reduces the probability and frequency of open-water-site usage (crowding-out models) and compared the predictive performance with that of models without crowding-out effects. For the

probability of open-water-site usage with crowding out, the model took the following form:

$$\begin{aligned} P(\text{water site usage}_{i,k} = 1) \\ = b_0 \times f_j(d_{i,k}) \times (1 - \alpha \times P(\text{tap or borehole usage}_i)) \end{aligned} \quad (5)$$

where the probability of open-water-site usage is inversely related to the probability of tap or borehole usage and where α quantifies the magnitude of crowding out. Again, this probability is estimated for each individual i and open-water site k pair. A model analogous to equation (5) is used for the duration of open-water-site usage.

Supplementary Table 7 provides a summary of all of the relevant spatial decay functions and models presented in this study.

Estimating spatial decay models

We used Bayesian nonlinear regression models, implemented using the `brms` package⁶³ in R version 4.1.0, to fit spatial decay models. Bayesian modelling was used because it allowed us to estimate distributions of the decay parameters and quantify the uncertainty of these parameter estimates by taking posterior draws from these distributions to obtain 95% CrIs. Due to convergence issues in the Bayesian modelling, we restricted our data to all sites within 3,000 m of the households, which captured over 99% of all open-water sites and taps or boreholes used. For all Bayesian models estimated via `brms`, a Hamiltonian Monte Carlo sampling algorithm via the Stan backend, employing the No-U-Turn Sampler, was used^{63,64}. Four Markov chains with 2,000 iterations per chain (1,000 warm-up iterations and 1,000 sampling iterations) were employed. Convergence was assessed based on whether the potential scale reduction factor (\hat{R}) was close to 1. Within `brms`, the `plot` function was used to assess chain mixing visually and the `pp_check` function was used to assess agreement between predicted and observed response values.

As far as possible, we used non-informative priors. For instance, in the exponential decay model, we used uniform priors Unif(0, 1) for b_0 and b_1 , the most conservative priors possible, as b_0 and b_1 can only take values between 0 and 1. Model fitting enforced $P(\text{response}_{i,j,k} = 1) \in [0, 1]$ throughout. Where model outputs were used to assign individuals to open-water sites, predicted probabilities were converted to binary outcomes using the optimal cut-point probability determined by the `cutpointR` package with the F1 score as the criterion⁶⁵.

Selecting the best-fitting models

To determine which decay function best fit the data, we used Bayesian leave-one-out cross-validation. The best fit was identified based on ELPD for a new dataset—a Bayesian leave-one-out measure of out-of-sample predictive fit⁶¹. When comparing a model against a reference model, negative ELPD values signify lower predictive density, indicating that a model performs worse than the reference model. For the decay function, ELPD was compared between global Hill, exponential and power-law decay models (Supplementary Table 7). We additionally report AUROCs and their 95% CIs from fivefold cross-validation to evaluate the ability of the models to correctly discriminate between used and unused open-water sites or taps or boreholes for each individual. When computing AUROCs, stratification based on district was performed when the folds were constructed to ensure proper representation of each district. In models where one district at a time was held out to evaluate predictive performance on data from this district using a model trained on the other two districts (Supplementary Fig. 12), the training data comprised 50 open-water sites per individual, sampled with replacement from the full data to ensure a more balanced dataset across districts when calculating AUROC.

Importantly, we evaluated the predictive performance of the spatial decay models at the individual level, which is a substantially more stringent approach compared with how gravity models (which are akin to our spatial decay models) are conventionally evaluated.

As gravity models are not at the individual level, Pearson correlation coefficients or R^2 values between predicted and observed aggregate mobility flows are typically used for evaluation^{66–68}. For fine-scale predictions, gravity models have performed poorly, with R^2 values for predicting commuting behaviour in London of between 7–22%⁶⁷. Here, we are not primarily interested in measures of aggregate performance but in the more difficult problem of predicting individual-level usage patterns. This problem is more challenging because we aim to predict not whether any open-water contact occurs but the specific open-water site(s) that individuals visit. Consequently, if an individual visits a different water site than the one predicted by the model, this would not be counted as a classification mistake when the goal is to predict the aggregate number of individuals with open-water contact. In our case, however, this case would introduce two classification mistakes (one open-water site incorrectly predicted as not being visited and another being incorrectly predicted as visited).

Out-of-sample predictions of open-water contact

The ability of spatial decay models to predict population-wide patterns of open-water contact was demonstrated using the publicly available Global Google–Microsoft Open Buildings Dataset (V3), which has mapped over two billion structures globally⁴². Footprints of all buildings in a 10 km × 10 km area around the study villages in Pakwach were extracted. As there were few commercial or public buildings in this rural area, we assumed that each building represented a household.

To generate population-wide predictions of open-water contact, we calculated the centroid location of each building and input the centroid GPS coordinates into equation (3), using the district-specific decay model for Pakwach. This produced out-of-sample predictions of open-water-site usage for all households in the study villages in Pakwach.

Sensitivity and robustness analyses

Several sensitivity and robustness analyses were conducted to account for the potential for open-water contact misclassification, systematic differences in compliance across demographic groups, and differential logger failure across districts.

As most open-water contact events are brief (in the range of minutes), there is a risk of misclassification, for instance due to passing by an open-water site without having actual open-water contact. To evaluate the influence of such brief, potentially misclassified contact events, we successively increased the threshold below which we removed events (0–10 min). We then recalculated the spatial decay model predicting open-water-site usage, including only open-water contact events above these thresholds, to understand how the model coefficients b_0 and b_1 changed. This thresholding exercise also informed on the minimum GPS logger sampling frequency needed to reproduce the patterns observed with GPS data at two-minute intervals.

Comparisons were made between the demographic characteristics of individuals with sufficient GPS data (≥ 2 days), those with insufficient data (< 2 days or completely missing data) and the broader SchistoTrack study population. This analysis provided insights into potential selection biases arising from sampling or the distribution of loggers and participant non-compliance. Among those with GPS data, terciles of the number of GPS points per individual per day (calculated within each district to account for deteriorating battery life due to device reuse) were used as an indicator to examine differences in non-compliance with wear between demographic groups.

This study was conducted across three different districts; behaviour patterns from one district may not be generalizable. To address this, data from two districts at a time were used to predict water site usage in the held-out district using the global spatial decay model of open-water contact (equation (2)). The numbers of individuals and observation days were downsampled to determine the minimum numbers of individuals and observation days required to achieve good

model performance, which was assessed based on AUROCs. Although participants were required to contribute a minimum of two days of GPS logger data to be included in this analysis, we assessed for included participants whether even a single day of GPS logger data from all participants yielded model performance based on the AUROC.

Spatially explicit transmission model

We constructed a modular, spatially explicit IBM to simulate the dynamics of reinfection in our study population using Julia (version 1.11.3). We informed our model from existing IBMs, adapting snail dynamics from the SchiSTOP model²¹, incorporating them at the site level rather than aggregated and adapting human infection dynamics from the SCHISTOX model²⁰. Parameter values were derived from these two models or based on biological realism, but were not directly fitted to the data. Initial snail states were informed by Iacovidou et al.²⁶, with aggregated information for the water site clusters, and initial worm burdens in humans were informed by the EPG data of the 452 individuals who participated in the GPS logger study.

Spatial explicitness in the model was introduced using weighted bipartite networks (one for each district), where nodes represent households and open-water sites and edges were weighted according to the probability, P , of site usage, as calculated using the spatial decay function (equation (2)).

We considered four decision rules regarding site usage, where each rule represents an independent visitation decision per individual per time step:

- Visit the nearest open-water site with 100% probability.
- Visit the nearest open-water site with probability P .
- Each site is given a success (1) or failure (0) score based on Bernoulli trials with probability P . Individuals visit a single site with a success score chosen at random. If no successes are recorded, individuals stay home.
- Same as the previous rule, but individuals visit all sites that return a success score from the Bernoulli trials.

The first three rules only allow individuals to visit a maximum of one site per day, with the first two rules restricting visitation to their nearest site. All rules, except the first one, are informed by the spatial decay function.

We simulated 100 replicates for each decision rule for 365 days, considering the study population of 452 individuals. The aims were to investigate whether information from the spatial decay models was informative for predicting infection prevalence one year after treatment and to assess the site-specific visitation distributions from the different rules. Site visits were calculated by counting, for each replicate, how many times individuals visited a given site, then summarizing these counts across all replicates to obtain average visit frequencies per site. By taking the cumulative miracidial input at the site level, we computed the relative contamination for each site as a percentage of the most contaminated site within each district.

We performed sensitivity analysis on the aggregation parameter (as was also done in Graham et al.²⁰ when using the SCHISTOX model) by showing the predicted prevalence results using lower values. However, we believe a higher value is more consistent with the observed prevalence in our study area.

Inclusion

This study was co-designed and implemented by researchers based in Uganda and the United Kingdom. Ugandan investigators from the Division of Vector-Borne and Neglected Tropical Diseases at the Ministry of Health and local partners contributed to the study design, community engagement, field implementation, data curation and manuscript preparation, and all listed authors meet journal authorship criteria, with additional contributors acknowledged. Characterizing fine-scale open-water contact behaviours and identifying key transmission sites emerged as a local priority through local community engagement

activities as part of SchistoTrack, and the research results are shared with communities during annual sessions. Research ethics approval was obtained from the Oxford Tropical Research Ethics Committee, Vector Control Division Research Ethics Committee of the Uganda Ministry of Health and Uganda National Council for Science and Technology. All participants received remuneration (10,000 UGX, equivalent to ~US\$2.73 or one day of wages) upon enrolment into SchistoTrack and an equal amount upon completion of the GPS logger sub-study. Those with confirmed *S. mansoni* infection received praziquantel in line with Ugandan national guidelines. No biological specimens, cultural artefacts or associated traditional knowledge were exported outside the country. Training and capacity-building workshops for household survey data collection, environmental sampling and spatial mapping using geographical information software were conducted as part of SchistoTrack.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Due to the highly sensitive nature of the GPS logger data, restrictions from the data protection impact assessment and the ongoing nature of the SchistoTrack cohort, these data are not publicly available. Researchers wishing to access the data may submit a written request to the corresponding author. We will consider requests from researchers with a legitimate scientific purpose, subject to the execution of a data sharing agreement with the University of Oxford that specifies permitted uses and any applicable data protection conditions. Extensive metadata have been provided in the manuscript and supplementary materials.

Code availability

The computer code used to produce all of the figures and tables in the main manuscript and supplementary materials is available on FigShare at <https://doi.org/10.25446/oxford.31573027> (ref. 69).

References

- Schistosomiasis. WHO <https://www.who.int/news-room/fact-sheets/detail/schistosomiasis> (2026).
- Reitzug, F. et al. Current water contact and *Schistosoma mansoni* infection have distinct determinants: a data-driven population-based study in rural Uganda. *Nat. Commun.* **15**, 9530 (2024).
- Pinot de Moira, A. et al. Analysis of complex patterns of human exposure and immunity to *Schistosomiasis mansoni*: the influence of age, sex, ethnicity and IgE. *PLoS Negl. Trop. Dis.* **4**, e820 (2010).
- Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases 2021–2030* (WHO, 2020); <https://www.iapb.org/wp-content/uploads/2021/02/A-roadmapfor-neglected-tropical-diseases-2021-30-eng.pdf>
- Mwinzi, P. N. et al. Priority knowledge gaps for schistosomiasis research and development in the World Health Organization Africa Region. *Infect. Dis. Poverty* **14**, 19 (2025).
- Musuva, R. M. et al. Unprotected water sources and low latrine coverage are contributing factors to persistent hotspots for schistosomiasis in western Kenya. *PLoS ONE* **16**, e0253115 (2021).
- Bethony, J. et al. Exposure to *Schistosoma mansoni* infection in a rural area in Brazil. Part III: household aggregation of water-contact behaviour. *Trop. Med. Int. Health* **9**, 381–389 (2004).
- Lamberti, O., Kabatereine, N. B., Tukahebwa, E. M. & Chami, G. F. *Schistosoma mansoni* infection risk for school-aged children clusters within households and is modified by distance to freshwater bodies. *PLoS ONE* **16**, e0258915 (2021).
- Lund, A. J. et al. Exposure, hazard, and vulnerability all contribute to *Schistosoma haematobium* re-infection in northern Senegal. *PLOS Negl. Trop. Dis.* **15**, e0009806 (2021).
- Ciddio, M. et al. The spatial spread of schistosomiasis: a multidimensional network model applied to Saint-Louis region, Senegal. *Adv. Water Resour.* **108**, 406–415 (2017).
- Hardwick, R. J., Vegvari, C., Collyer, B., Truscott, J. E. & Anderson, R. M. Spatial scales in human movement between reservoirs of infection. *J. Theor. Biol.* **524**, 110726 (2021).
- Mari, L. et al. Big-data-driven modeling unveils country-wide drivers of endemic schistosomiasis. *Sci. Rep.* **7**, 489 (2017).
- WHO *Guideline on Control and Elimination of Human Schistosomiasis* (WHO, 2022); <https://apps.who.int/iris/handle/10665/351856>
- Grimes, J. E. T. et al. The relationship between water, sanitation and schistosomiasis: a systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* **8**, e3296 (2014).
- Knopp, S. et al. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. *Lancet Glob. Health* **7**, e1118–e1129 (2019).
- Reitzug, F., Ledien, J. & Chami, G. F. Associations of water contact frequency, duration, and activities with schistosome infection risk: a systematic review and meta-analysis. *PLoS Negl. Trop. Dis.* **17**, e0011377 (2023).
- Arnold, B. F. et al. Fine-scale heterogeneity in *Schistosoma mansoni* force of infection measured through antibody response. *Proc. Natl Acad. Sci. USA* **117**, 23174–23181 (2020).
- Exum, N. G. et al. The prevalence of schistosomiasis in Uganda: a nationally representative population estimate to inform control programs and water and sanitation interventions. *PLoS Negl. Trop. Dis.* **13**, e0007617 (2019).
- Perez-Saez, J. et al. A theoretical analysis of the geography of schistosomiasis in Burkina Faso highlights the roles of human mobility and water resources development in disease transmission. *PLOS Negl. Trop. Dis.* **9**, e0004127 (2015).
- Graham, M. et al. SCHISTOX: an individual based model for the epidemiology and control of schistosomiasis. *Infect. Dis. Model.* **6**, 438–447 (2021).
- Malizia, V., de Vlas, S. J., Roes, K. C. & Giardina, F. Revisiting the impact of *Schistosoma mansoni* regulating mechanisms on transmission dynamics using SchiSTOP, a novel modelling framework. *PLoS Negl. Trop. Dis.* **18**, e0012464 (2024).
- Sow, S. et al. The contribution of water contact behavior to the high *Schistosoma mansoni* infection rates observed in the Senegal River Basin. *BMC Infect. Dis.* **11**, 198 (2011).
- Kloos, H. et al. Spatial patterns of human water contact and *Schistosoma mansoni* transmission and infection in four rural areas in Machakos District, Kenya. *Soc. Sci. Med.* **44**, 949–968 (1997).
- Woolhouse, M. E. J. et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342 (1997).
- Puthur, C., Nabatte, B., Tinkitina, B., Kabatereine, N. B. & Chami, G. F. Design and maintenance of community-based cohorts in sub-Saharan Africa: a longitudinal evaluation of participant attrition in SchistoTrack. *BMJ Public Health* **3**, e002151 (2025).
- Iacovidou, M. et al. Temporal variability and flooding influence the ecological niche of *Biomphalaria* intermediate hosts for *Schistosoma mansoni* in rural Uganda. *Proc. R. Soc. B Biol. Sci.* **293**, 20252083 (2026).
- Eyre, M. T. et al. Piloting an integrated approach for estimation of environmental risk of *Schistosoma haematobium* infections in pre-school-aged children and their mothers at Barombi Kotto, Cameroon. *Acta Trop.* **212**, 105646 (2020).
- Ruiz Cuenca, P. et al. Using step selection functions to analyse human mobility using telemetry data in infectious disease epidemiology: a case study of leptospirosis. *eLife* **14**, RP107153 (2025).

29. Wesolowski, A., O'Meara, W. P., Eagle, N., Tatem, A. J. & Buckee, C. O. Evaluating spatial interaction models for regional mobility in sub-Saharan Africa. *PLoS Comput. Biol.* **11**, e1004267 (2015).
30. Meredith, H. R. et al. Characterizing human mobility patterns in rural settings of sub-Saharan Africa. *eLife* **10**, e68441 (2021).
31. Seto, E. Y. W. et al. Patterns of intestinal schistosomiasis among mothers and young children from Lake Albert, Uganda: water contact and social networks inferred from wearable global positioning system dataloggers. *Geospat. Health* **7**, 1–13 (2012).
32. Cairncross, S. Trachoma and water. *Community Eye Health* **12**, 58 (1999).
33. Woolhouse, M. E., Watts, C. H. & Chandiwana, S. K. Heterogeneities in transmission rates and the epidemiology of schistosome infection. *Proc. R. Soc. Lond. B Biol. Sci.* **245**, 109–114 (1991).
34. Zheng, L. et al. Molluscicides against the snail-intermediate host of *Schistosoma*: a review. *Parasitol. Res.* **120**, 3355–3393 (2021).
35. Proud, R. et al. Stocking African catfish in Lake Victoria provides effective biocontrol of snail vectors of *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* **19**, e0013490 (2025).
36. Trippler, L. et al. Test–treat–track–test–treat (5T) approach for *Schistosoma haematobium* elimination on Pemba Island, Tanzania. *BMC Infect. Dis.* **24**, 661 (2024).
37. Jordan, P. Epidemiology and control of schistosomiasis. *Br. Med. Bull.* **28**, 55–59 (1972).
38. Sangare, M. et al. Understanding the barriers and facilitators related to never treatment during mass drug administration among mobile and migrant populations in Mali: a qualitative exploratory study. *BMJ Glob. Health* **9**, e015671 (2024).
39. Adams, M. W. et al. Leaving no one behind: targeting mobile and migrant populations with health interventions for disease elimination—a descriptive systematic review. *BMC Med.* **20**, 172 (2022).
40. Parker, M. et al. Border parasites: schistosomiasis control among Uganda's fisherfolk. *J. East. Afr. Stud.* **6**, 98–123 (2012).
41. *Field Use of Molluscicides in Schistosomiasis Control Programmes: an Operational Manual for Programme Managers* (WHO, 2017); <https://iris.who.int/server/api/core/bitstreams/2f575968-491e-4e89-918c-4c88857f8235/content>
42. Google-Microsoft Open Buildings - combined by VIDA. *Source Cooperative* <https://source.coop/vida/google-microsoft-open-buildings> (2023).
43. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data* **4**, 170004 (2017).
44. *Joint Monitoring Programme* (WHO & UNICEF, accessed 20 November 2025); <https://washdata.org/data>
45. Deol, A. K. et al. Schistosomiasis—assessing progress toward the 2020 and 2025 global goals. *N. Engl. J. Med.* **381**, 2519–2528 (2019).
46. Sedgwick, P. The Hawthorne effect. *BMJ* **344**, d8262 (2011).
47. Calabrese, J. M., Fleming, C. H. & Gurarie, E. ctmm: an r package for analyzing animal relocation data as a continuous-time stochastic process. *Methods Ecol. Evol.* **7**, 1124–1132 (2016).
48. Morris, G. & Conner, L. M. Assessment of accuracy, fix success rate, and use of estimated horizontal position error (EHPE) to filter inaccurate data collected by a common commercially available GPS logger. *PLoS ONE* **12**, e0189020 (2017).
49. Berger-Tal, O. & Bar-David, S. Recursive movement patterns: review and synthesis across species. *Ecosphere* **6**, 1–12 (2015).
50. Seidel, D. P., Dougherty, E., Carlson, C. & Getz, W. M. Ecological metrics and methods for GPS movement data. *Int. J. Geogr. Inf. Sci.* **2**, 2272–2293 (2018).
51. Bracis, C. recurse: Computes revisitation metrics for trajectory data. R version 1.4.0 <https://CRAN.R-project.org/package=recurse> (2024).
52. Hahsler, M., Piekenbrock, M. & Doran, D. dbSCAN: fast density-based clustering with R. *J. Stat. Softw.* **91**, 1–30 (2019).
53. Kwiringira, J., Atekyereza, P., Niwagaba, C. & Günther, I. Gender variations in access, choice to use and cleaning of shared latrines; experiences from Kampala Slums, Uganda. *BMC Public Health* **14**, 1180 (2014).
54. Tessema, R. A. Assessment of the implementation of community-led total sanitation, hygiene, and associated factors in Diretiyara district, Eastern Ethiopia. *PLoS ONE* **12**, e0175233 (2017).
55. Webster, B. L. et al. Praziquantel treatment of school children from single and mixed infection foci of intestinal and urogenital schistosomiasis along the Senegal River Basin: monitoring treatment success and re-infection patterns. *Acta Trop.* **128**, 292–302 (2013).
56. Tchuem Tchuenté, L. A., Momo, S. C., Stothard, J. R. & Rollinson, D. Efficacy of praziquantel and reinfection patterns in single and mixed infection foci for intestinal and urogenital schistosomiasis in Cameroon. *Acta Trop.* **128**, 275–283 (2013).
57. Buuren, S. V. & Groothuis-Oudshoorn, K. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).
58. Simini, F., González, M. C., Maritan, A. & Barabási, A. L. A universal model for mobility and migration patterns. *Nature* **484**, 96–100 (2012).
59. González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
60. Pappalardo, L. et al. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166 (2015).
61. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2016).
62. Stouffer, S. A. Intervening opportunities: a theory relating mobility and distance. *Am. Sociol. Rev.* **5**, 845–867 (1940).
63. Bürkner, P.-C. brms: an R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
64. Hoffman, M. D. & Gelman, A. The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
65. Thiele, C. & Hirschfeld, G. cutpointR: improved estimation and validation of optimal cutpoints in R. *J. Stat. Softw.* **98**, 1–27 (2021).
66. Yang, Y., Herrera, C., Eagle, N. & González, M. C. Limits of predictability in commuting flows in the absence of data for calibration. *Sci. Rep.* **4**, 5662 (2014).
67. Masucci, A. P., Serras, J., Johansson, A. & Batty, M. Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. *Phys. Rev. E* **88**, 022812 (2013).
68. Li, W., Wang, Q., Liu, Y., Small, M. L. & Gao, J. A spatiotemporal decay model of human mobility when facing large-scale crises. *Proc. Natl Acad. Sci. USA* **119**, e2203042119 (2022).
69. Reitzug, F. et al. Predicting schistosome transmission in rural Uganda using water contact data from wearable GPS devices. *Figshare* <https://doi.org/10.25446/oxford.31573027> (2026).
70. *Uganda GIS Data* (World Resource Institute, accessed 17 March 2026); <https://www.wri.org/data/uganda-gis-data#base>

Acknowledgements

A PhD scholarship was awarded from the Nuffield Department of Population Health to F.R. Grants from the Wellcome Trust Institutional Strategic Support Fund (204826/Z/16/Z) and Nuffield Department of Population Health Pump Priming Fund and a Robertson Foundation Fellowship were awarded to G.F.C. We thank all study participants and local communities for their involvement in the study. We also thank all field teams, including malacologists, surveyors, nurses and laboratory technicians involved in the 2022 SchistoTrack baseline data collection. The support of the Uganda Ministry of Health, local district leaders, focal health workers and village health teams was crucial for

building partnerships and continued trust with study communities. We thank F. L. Shabaan for help with preparing the GPS loggers for fieldwork. After the fieldwork, A. Fowler helped with initial attempts to retrieve data from damaged GPS loggers. L. B. Wilburn, K. Foster and J. Brown pulled GPS logger data from damaged GPS loggers. P. Sambaturu gave feedback during the initial stages of the analysis. We also thank M. M. Lang for statistical insights.

Author contributions

F.R. and G.F.C. conceived of the article idea. F.R., M.A.I., N.B.K., B.N., M.T.E., B.T., A.M.B. and G.F.C. curated the data. F.R. and M.A.I. performed the formal analysis, validated the results, visualized the results and wrote the original draft of the manuscript. G.F.C. acquired funding, developed the software and supervised the study. F.R., M.A.I. and G.F.C. performed the investigation and developed the methodology. N.B.K. and G.F.C. were responsible for project administration and provided resources. F.R., M.A.I., N.B.K., B.N., M.T.E., B.T., A.M.B., R.L. and G.F.C. reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s44360-026-00118-w>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44360-026-00118-w>.

Correspondence and requests for materials should be addressed to Goylette F. Chami.

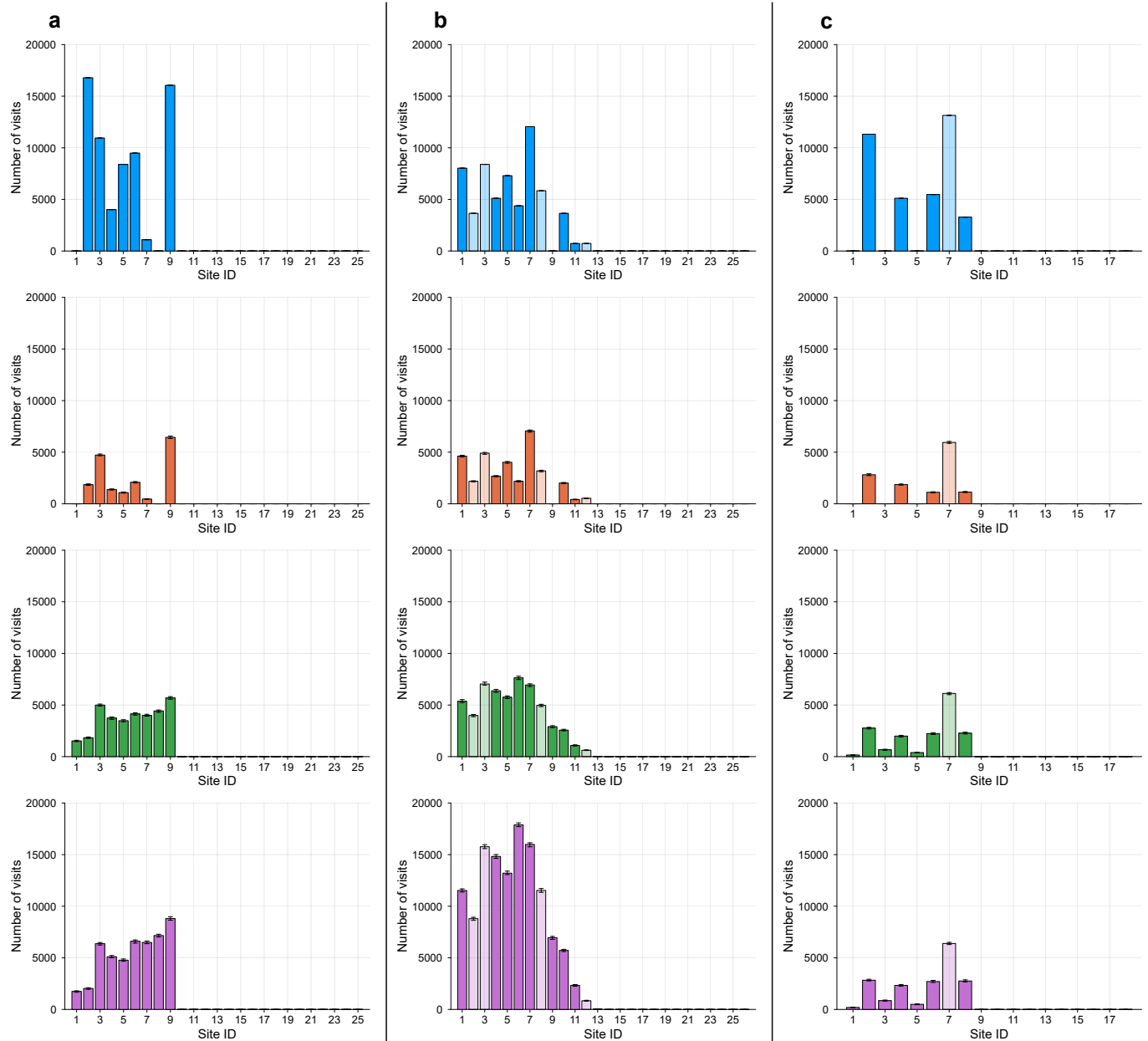
Peer review information *Nature Health* thanks David Musoke, Javier Perez Saez, W. Evan Secor and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Lorenzo Righetto, in collaboration with the *Nature Health* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026



Extended Data Fig. 1 | Predicted total number of visits per open-water site over 365 days for each district and visitation rule. The panels show the median predicted total number of open water site visits over 365 days and 100 replicates across the districts of Pakwach (a), Buliisa (b), and Mayuge (c) under the four different visitation rules (blue-rule (1), orange-rule (2), green-rule (3), and purple-rule (4)). Each vertical bar represents one open water site with the x-axis label

denoting its site ID. Lighter-coloured bars correspond to open water sites where no snails were observed during snail sampling, and which consequently do not contribute to transmission. Error bars represent the 2.5th and 97.5th percentile range across replicates. The narrow intervals indicate low between-replicate variability in predicted site visitation. Note that rule (1) is deterministic, so between-replicate variability for this rule is zero, and no error bars are shown.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Due to the highly sensitive nature of the GPS logger data, restrictions from the data protection impact assessment, and ongoing nature of the SchistoTrack cohort, these data are not publicly available. Extensive metadata for all variables have been provided in the manuscript and supplement.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We exclusively use the term gender, following the definitions laid out in the 'Sex and Gender Equity in Research – SAGER – guidelines' because in this study, information was elicited via self-reports and we sought to understand how gender (shaped by social and cultural norms) influences water contact behaviour.
Reporting on race, ethnicity, or other socially relevant groupings	We recorded data on self-reported occupation, gender, and age
Population characteristics	See behavioral & social study design section below
Recruitment	For the SchistoTrack baseline, we randomly sampled a total of 1,459 households (n=2885)—approximately 40 per village—in 38 villages across three districts, using village registers or mass drug administration (MDA) records. All households with at least one child and one adult residing in the village for at least six months of the year were eligible. For the GPS logger study presented in this paper, we subsequently sampled four villages per district (12 in total) based on their levels of water contact and occupational fishing, determined predominantly by the presence of a beach or landing site. None of the selected villages had piped water. A maximum of 50 participants per village were recruited, with a target of one adult-child pair (an adult ≥ 18 years] and one child [5–17 years]) per household.
Ethics oversight	Data collection and use were reviewed and approved by Oxford Tropical Research Ethics Committee (OxTREC 509-21), Vector Control Division Research Ethics Committee of the Uganda Ministry of Health (VCDREC146), and Uganda National Council of Science and Technology (UNCST HS 1664ES).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Cross-sectional, quantitative study of 452 individuals aged 5-82 years in Eastern and Western Uganda. We collected socio-demographics, biomedical information (including schistosome infection information), WASH information, environmental data, self-reported water contact data, and GPS-logger derived water contact, tap/borehole usage and human movement.
Research sample	Subsample of n=452 was drawn from a population-representative sample of n=2885. Participants were aged 5-82 years and 50% (226/452) were male and 50% (226/452) were female.
Sampling strategy	We selected 452 participants from the larger SchistoTrack study (n=2885) which had been randomly selected from village registers or MDA records. For the GPS logger study, participants were sampled from the SchistoTrack study study participant list to ensure a balance of age and gender, and fishermen were oversampled to ensure a diverse representation of open water contact and mobility patterns. We excluded participants that were physically immobile and thus unable to have water contact and we also excluded children who were sent away to boarding school. All participants were unaware of their infection status when they entered the GPS logger study.
Data collection	Household data collection was done by trained surveyors (either from Kampala or locally recruited) using tables and ODK software. Household-level information was collected prior to clinical data collection and participants were unaware of their infection status. All clinical assessments were conducted by technicians/nurses from the district or Kampala in mobile in-the-field labs. Wearable GPS loggers (i-gotU GT-600; Mobile Action Technology, Taiwan) were used to record the location data of participants at two-minute intervals between 5 a.m. and 8 p.m. local time over approximately ten consecutive days.
Timing	Data collection took place in January to February 2022 in Pakwach, Buliisa and Mayuge.
Data exclusions	From the 585 participants that were originally recruited were recruited for the study 133 were excluded from the analysis, mainly to device failure and loggers not being returned by participants after being handed out in Pakwach. After excluding participants with fewer than two complete days of GPS data. A flow chart detailing any data exclusions is provided in Fig S18.
Non-participation	One participants that was selected for the GPS logger study refused (see Flowchart Fig S18).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern
- Plants

Methods

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study protocol

Data collection

Outcomes

Plants

Seed stocks

Novel plant genotypes

Authentication