Local habitual movement as a mechanism for *Schistosoma mansoni* transmission resurgence – a causal analysis

Rivka M. Lim[1,2]*, Moses Arinaitwe[2,3], Simon A. Babayan[2], Andrina Nankasi[3], Alon Atuhaire[3], Annet Namukuta[3], Mwima Nicholas [3], Amy B. Pedersen[1], Joanne P. Webster[4], Poppy H.L. Lamberton[2]•, Jessica Clark[2]•

[1]Institute of Evolution and Ecology, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK

[2]School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, UK

[3]Vector Borne and Neglected Tropical Diseases Control Division, Ministry of Health, Kampala, Uganda

[4]Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, Herts, UK

*corresponding author: rivka.lim@ed.ac.uk;

•joint senior authorship

Abstract

Background/aims

The World Health Organization (WHO) aims to eliminate schistosomiasis as a public health problem (EPHP) across 78 endemic countries by 2030. However, for low-prevalence settings that reach EPHP, guidance on managing transmission to maintain EPHP or move towards Interruption of Transmission (IoT) is limited, partly due to insufficient evidence on drivers of resurgence. In Uganda, some communities inland from Lake Victoria have achieved EPHP for *Schistosoma mansoni* but not progressed to IoT. This study explored whether routine, short-range travel to the highly endemic lake could sustain transmission in these settings.

Methods

We conducted a cross-sectional study in five Ugandan villages ~5 km from Lake Victoria. Parasitological data were collected using Kato-Katz and Point-of-Care Circulating Cathodic Antigen tests, alongside questionnaires on lake travel from 585 individuals aged 1–91 years. A structural causal model estimated the total and direct effects of travel frequency, activity type, water contact duration, and drug treatment history on infection. Bayesian regression models and counterfactual simulations predicted infection under hypothetical interventions.

Results

Reaching IoT in low-risk villages may be undermined by habitual, short-range travel to high-risk sites, driven by the nature and duration of lake contact. Daily lake travel caused a 1.7-fold increase in odds of infection, while occupational activities caused a 3.4-fold increase compared with no lake activity. Counterfactual analysis showed that removing lake contact duration most reduced infection risk among moderate-frequency travellers, while daily travellers showed smaller changes, and some transmission persisted among individuals with little or no lake contact. Simulations demonstrated that modifying lake contact behaviours could reduce individual infection risk but had limited population-level impact.

Conclusion

These findings indicate that targeting only high-risk villages or individual behaviours is unlikely to achieve sustained, wide-spread IoT, underscoring the need for integrated control strategies that account for mobility, behaviour, and local transmission ecology.

<u>Key questions</u>

What is already known on this topic?

- There is considerable guidance provided by the WHO for achieving EPHP, based on a wealth of scientific evidence. However, evidence to support the development of resources for areas that have reached EPHP is limited, particularly with regards to when and how to scale back mass drug administration (MDA), whilst mitigating the risk of resurgence.
- Simulation studies suggest that reducing treatment in low-prevalence settings may trigger a return to higher transmission, but the mechanisms sustaining residual infections in these contexts remain unclear and are often identified through statistical associations rather than through robust causal analyses. As such, there is a lack of evidence identifying drivers of resurgence.
- Human mobility is recognised as a potential driver of schistosomiasis transmission. However, previous studies have focused on long-distance migration or spatial risk modelling to predict future impacts of movement. The impact of individual-level frequent, habitual, local travel between low- and high-prevalence areas on transmission and resurgence has received comparatively less attention.

What this study adds?

- This study provides causal evidence, rather than just statistical associations, that daily travel from low-prevalence inland villages to high-prevalence lakeshore landing sites increases the probability of *S. mansoni* infection risk and intensity, largely through activity type and duration of water contact.
- Counterfactual and mechanistic simulations show that reducing such travel would lower infection risk, but transmission persists among individuals with little or no lake

contact, suggesting that without MDA, travellers may sustain infection in their home villages.

How might this study affect research, practice or policy?

- Our findings suggest that low-prevalence communities should not automatically receive less frequent MDA.
- Equally, nor should communities who have achieved EPHP see their MDA frequency reduced without understanding sources of and plans for mitigating re-introduction of infection.
- Surveillance in such settings should attempt to capture mobile individuals and use more sensitive diagnostics to avoid underestimating residual risk.
- That communities are not isolated in space should be recognised when developing policy guidance, due to the risk of infection resurgence.
- Research and policy should continue to leverage causal approaches to clarify the mechanisms sustaining transmission in low-risk communities.

134 **Introduction**

135

136 Schistosomiasis is a parasitic disease caused by trematode worms of the genus *Schistosoma*.
137 In sub-Saharan African, where over 90% of human cases occur, the dominant species are *S.*
138 *mansoni* and *S. haematobium* (causing intestinal and urogenital disease, respectively).
139 Transmission occurs through contact with freshwater that has been contaminated through
140 open defecation or urination, or inadequate containment of urine or stool infected with
141 parasite eggs. These eggs then hatch and infect intermediate-host freshwater snails, where
142 they develop into larvae which are infective to humans when released from the snails.
143 Schistosomiasis transmission is highly localised, with infection prevalence and intensity
144 varying substantially even between neighbouring communities (1). In response to the
145 significant health burden posed by schistosomiasis, the World Health Organization (WHO)
146 recommends a coordinated set of interventions, including primarily mass drug
147 administration (MDA) with praziquantel, in combination with health education, improved
148 access to safe water and sanitation, and snail control (2,3). While these strategies have
149 reduced prevalence in many settings, it remains unclear whether, when and how best to
150 scale back MDA, with the 2021-2030 Neglected Tropical Disease Roadmap, developed by
151 the WHO, highlighting the need to understand how to monitor for, or what mechanisms
152 could drive resurgence (4).

153

154 The WHO outlines three sequential programmatic goals to guide control efforts against
155 schistosomiasis: 1) morbidity control; 2) elimination as a public health problem (EPHP); and
156 3) interruption of transmission (IoT) (2). An area is defined as having achieved these goals by
157 using infection intensity, which in the case of intestinal schistosomiasis is quantified by egg
158 counts generated using the Kato-Katz technique (5). For intestinal infection, a community is
159 considered to have achieved morbidity control when <5% of school aged children (SAC) are
160 heavily infected (defined as ≥400 eggs per gram of faeces (EPG)), EPHP is considered to have
161 been achieved when <1% of SAC are heavily infected, and IoT is defined as having no
162 autochthonous human or snail cases for five consecutive years (2). Progress towards these
163 targets is assessed through combined surveillance, monitoring egg-based infection intensity

and prevalence (5), and in some cases supported by Point-of-Care Circulating Cathodic Antigen tests (POC-CCAs) (6).

Whilst the WHO provides clear guidance for disease management and achieving EPHP in moderate and high prevalence settings (2), there are limited resources EPHP has been achieved. However, simulation models have indicated that when prevalence is <1% (by Kato Katz), residual persistent transmission can lead to resurgence without sufficient treatment (7) and surveillance (8). Furthermore, the impact of this is evidenced by resurgence in China following the reduction of interventions (9) and in Zanzibar after a one-year pause in MDA (10). Despite this risk, there are currently no robust post-EPHP surveillance and monitoring protocols (11), nor defined indicators for when to stop MDA or surveillance. Without clear guidance on how to detect early signs of resurgence or respond effectively, areas approaching EPHP or IoT, risk losing any progress made. To mitigate this risk, it is essential to identify the mechanisms that sustain low-level transmission and may drive resurgence, to inform surveillance and treatment strategies critical to long-term elimination and interruption success.

Communities living on the shores of Lake Victoria, Uganda tend to rely heavily on the lake for daily life. However, despite two decades of MDA, many continue to experience high rates of *S. mansoni* transmission (12,13). In contrast, some inland communities have shown reductions in prevalence, with some reaching EPHP, albeit without achieving IoT (14). This raises the question of whether inland communities which have reached EPHP, fail to progress to IoT due to their proximity to high-endemicity regions, and specifically whether transmission is sustained through habitual human movement between these low- and high-risk locations.

Local habitual human mobility between low and high transmission regions has received comparably less attention than long-term/distance human migration or movement (15–18). Where previous studies have identified human mobility as a key factor shaping the spatial dynamics of schistosomiasis transmission, they have largely relied on indirect evidence, such as spatial models of risk, or ecological correlations, to infer associations between movement and infection (19–21) and have not focussed on the impact of habitual movement between

areas which have achieved EPHP and those where endemicity remains high. Existing control strategies often operate by administrative units, assuming that populations are geographically isolated, and as such likely obscuring the possibility that habitual local mobility could mean individuals act as a conduit for sustained transmission. This raises another important question of whether certain people should, or can, be prioritised for surveillance or treatment during the elimination phase, as a function of their connectivity to high-risk locations, to prevent re-establishment of local transmission.

Additionally, although travel-related behaviours such as water contact and time spent in the lake are known risk factors for schistosomiasis transmission (22), there is little evidence on whether particular individuals, or their activities, are the cause of this risk. Understanding whether routine travel directly increases infection risk, or whether it acts indirectly through other behaviours is essential to inform targeted elimination strategies. Traditional epidemiological analyses often describe statistically significant associations between exposures and outcomes but cannot determine whether those exposures drive the outcomes (i.e. evidence causality). In contrast, causal inference methods, like structural causal modelling (SCM), are specifically used to estimate the effects of exposures while accounting for confounding, mediation (where effects operate through intermediate variables), and complex interdependencies. These approaches are increasingly applied in infectious disease epidemiology to generate policy-relevant insights, where programmatic decisions must be made using observational data (23,24). By explicitly defining and testing assumptions regarding the ways different drivers might interact, and the pathways through which transmission is shaped, SCMs offer a rigorous framework to evaluate drivers of infection and support elimination planning.

The aim of this study was to quantify the causal effect of habitual, local travel to and from highly endemic Lake Victoria areas on *S. mansoni* infection risk in regions that have met WHO targets for EPHP. To do this, we applied a Bayesian causal inference modelling framework to highly detailed behavioural and parasitological data collected from five communities, to estimate the direct and indirect effects of travel-related behaviours on infection risk. Understanding these patterns of movement and their relationship to infection

227 can offer valuable insight into whether such individuals may be contributing to ongoing
228 transmission that could be hindering progress toward IoT. Clarifying this relationship will
229 support the development of efficient and effective surveillance strategies, improve disease
230 resurgence detection, and ultimately contribute to achieving IoT goals.
231

232 **Methods**

233

234 Ethics

235 Informed written consent via signature or thumbprint was gathered prior to any data
236 collection for each adult, or from a legal guardian of those under the age of 18. In addition
237 to guardian consent, those aged eight to 17 gave informed signed or thumb printed assent.
238 All participants understood that they could leave the study at any point without losing
239 access to treatment. Enrolled participants were allocated a numerical ID which pseudo-
240 anonymously linked parasitological and travel data.
241

242 Public involvement

243 Community members were first engaged during a sensitisation visit led by Ministry of
244 Health team. The research questions were developed from policy priorities to understand
245 community experiences and practices related to schistosomiasis and its treatment,
246 particularly in low-risk areas where data are limited. While the public did not directly inform
247 study design, outcome measures, or recruitment, participation was voluntary following the
248 sensitisation meetings, during which questions were addressed before consent. Feasibility
249 was thus dependent on community acceptance. The pilot questionnaire was refined based
250 on volunteer feedback.
251

252 Data collection

253 Five villages each ~5 km from Lake Victoria, Uganda were selected: B, Nam, Nan, Nw and S
254 (Figure 1) (see supplementary data "Village selection" for full selection protocol). Individuals
255 aged one-91 years old were recruited assisted by the village health team. The study took
256 place in each village for one week, from October 2021- December 2021.
257

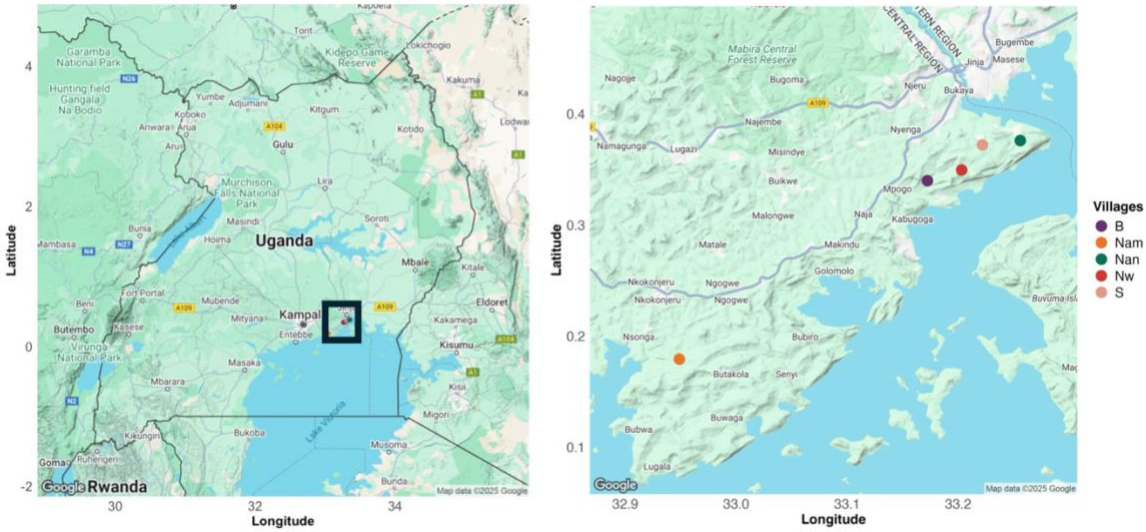


Figure 1. Map of the study area showing the locations of the five participating villages along the northern shore of Lake Victoria, Uganda. Each point represents one village, colour-coded by name: B (purple), Nw (red), Nam (orange), Nan A (green), and S (pink). The left panel provides a broader map of Uganda, with a black rectangle indicating the location of the zoomed-in map shown in the right panel, which displays a detailed view of the study area.

Three stool samples were collected per person over three days and duplicate Kato-Katz examined by experienced technicians, under a microscope, for *S. mansoni* eggs (5). One urine sample per person was tested using POC-CCAs, using the 1-10 G-Score system (25). Kato-Katz have 100% specificity but POC-CCAs were included to improve sensitivity in low prevalence settings (26). A participant was considered positive for *S. mansoni* if at least one egg was observed in a stool sample, or if negative by eggs but a POC-CCA ≥G3 (27,28). Prevalence was calculated by dividing the number of positive people by the full sample size. Individual mean infection intensity (EPG) was calculated as the mean slide count multiplied by 24. Population-level mean EPG was calculated removing individuals with EPG=0.

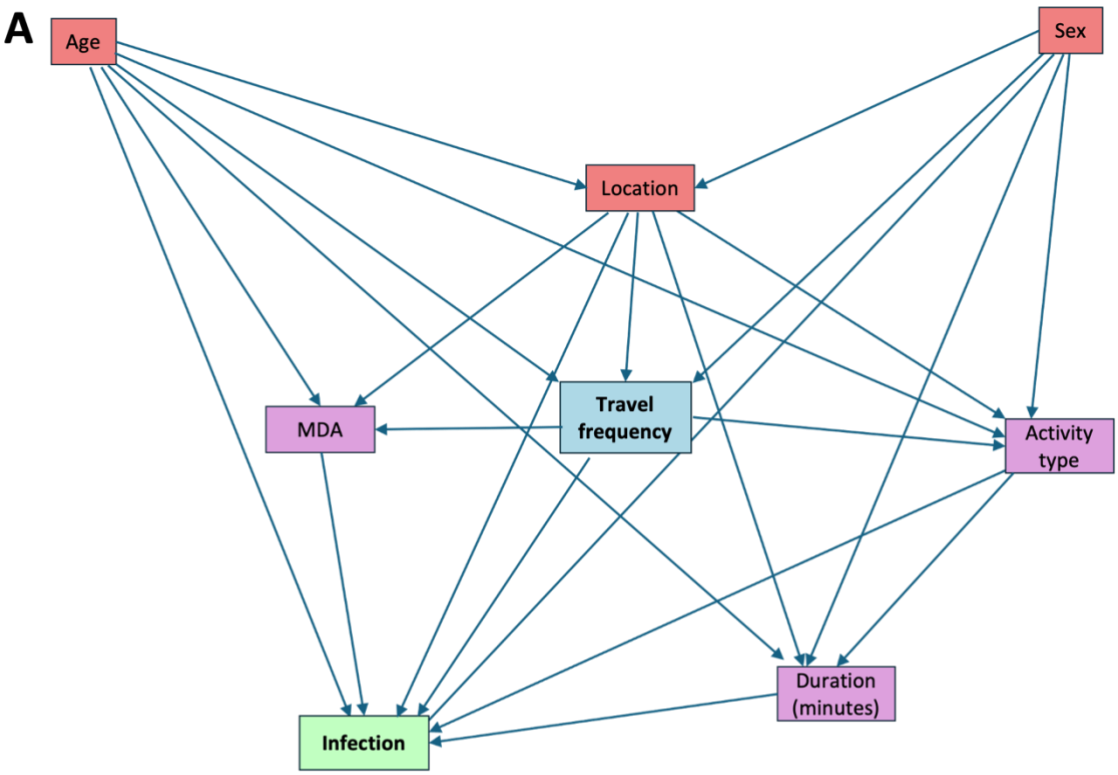
Human mobility and water contact were assessed through a questionnaire administered once during the study period. Questionnaires were conducted in the local language by trained fieldworkers and answers recorded on tablets with KoboToolBox. Respondents were asked their age, sex, and current village, as well as how many times they had travelled to

Lake Victoria in the past three months: i) never, ii) just once, iii) once a month, iv) twice a month, v) once a week, vi) twice a week, or vii) daily; what activities they engaged in upon arrival and how long they were in the water. Age was categorised into three programmatically relevant groups: PSAC (0–5 years), SAC (6–15 years), and adults (>15 years). We explored an alternative four-level specification including young adults (15-25 years) in sensitivity analyses. As effect estimates for young adults were comparable to those for older adults, with no improvement in model fit or predictive accuracy, the three-level age variable was retained for the analysis. Questions regarding MDA participation were also included.

Participants reported a wide variety of activities undertaken at the lake. These activities were grouped into i) domestic, ii) occupational, iii) recreational, and iv) trade/visit and were used in all models. A complete list of activity responses and their groupings is provided in Table S1.

Directed acyclic graph construction

All analyses and visualisations were executed using R (v. 4.4.3). A directed acyclic graph (DAG) was constructed to represent hypothesised causal relationships influencing the link between an individual's infection status and travel to a highly endemic location. The DAG was then used to inform the structure of the final model used for analysis. This DAG included travel frequency to Lake Victoria, water contact duration at Lake Victoria, activity type at the lake, MDA history, and demographic covariates: age, sex, and village of residence (Figure 2A). The structure of the DAG i.e. direction of the edges and which variables interact, was informed by a combination of domain knowledge, published literature, and expert opinion of schistosomiasis transmission dynamics, local environmental conditions, and behavioural patterns (29). The DAG aimed to visually represent a biologically plausible and theoretically grounded causal framework. The DAG was visualised using the R package dagitty (30). The structure of the DAG was validated against the observed data by testing the implied conditional independencies between the variables (see supplementary data "Model validation and restructuring" for a full explanation).



B

SCM:

$$\mathrm{Ag} := \mathcal{U}_{Ag}$$
$$\mathrm{S} := \mathcal{U}_{S}$$
$$\mathrm{L} := \beta_{Ag}\, Ag + \beta_{S}\, S + \mathcal{U}_{L}$$
$$\mathrm{M} := \beta_{Ag}\, Ag + \beta_{F}\, F + \beta_{L}\, L + \mathcal{U}_{M}$$
$$\mathrm{F} := \beta_{Ag}\, Ag + \beta_{S}\, S + \beta_{L}\, L + \mathcal{U}_{F}$$
$$\mathrm{A} := \beta_{F}\, F + \beta_{Ag}\, Ag + \beta_{S}\, S + \beta_{L}\, L + \mathcal{U}_{A}$$
$$\mathrm{D} := \beta_{Ag}\, Ag + \beta_{S}\, S + \beta_{A}\, A + \beta_{L}\, L + \mathcal{U}_{D}$$
$$\mathrm{I} := \beta_{Ag}\, Ag + \beta_{S}\, S + \beta_{A}\, A + \beta_{D}\, D + \beta_{F}\, F + \beta_{L}\, L + \beta_{M}\, M + \mathcal{U}_{F}$$

Figure 2 Directed acyclic graph (DAG) representing hypothesised causal relationships between variables and *Schistosoma mansoni* infection (A), with structural equations for each node (B). (A) The DAG outlines the assumed causal structure linking demographic, behavioural, and environmental variables to infection. Arrows represent direct causal effects. Key exposure variables include travel frequency, activity type, and duration of water contact, with potential confounders such as age class, sex, and location. (B) Structural

equations define the dependencies of each variable, with terms representing regression coefficients (β), unobserved residual error (U), and observed covariates. Ag = age class, S = sex, L = location, M = MDA, F = frequency of travel, A = activity, D = duration in the lake, I = infection.

Structural causal model

From the fully validated DAG, a SCM was constructed for analysis. Each variable was expressed as a function of its direct causes (parents in the DAG) and an independent noise term (Figure 2B). These functions define the joint distribution of all variables in the system, informing the underlying causal structure. Minimal adjustment sets for each exposure-outcome relationship were estimated from the SCM to ensure an unbiased estimation of total and direct effects (Table S2) (31,32). These adjustment sets were formulated using causal graph theory principles based on chains, forks and colliders (33). To account for uncertainty and provide probabilistic estimates of causal effects, we used Bayesian linear regression models. Each edge in the DAG was parameterised using these Bayesian models. Effects were considered supported if the 95% credible interval (95% crI) for the effect between variables did not include zero after adjusting for the conditioning set identified by the DAG structure. Using these appropriate adjustment sets we estimated total and direct effects of key causes of infection status and intensity, fitting models for each of the four primary causes: travel frequency, water contact duration, MDA history and activity type. Bayesian linear models were fitted, and posterior predictive checks conducted using the *brms* package in R (34). Binary infection status (*S. mansoni* presence or absence) was modelled using a Bernoulli distribution with a logit link. Egg burden (measured as mean EPG, among egg-positive individuals only) was modelled using a Gamma distribution with a log link, given its positive and right-skewed distribution. After carrying out sensitivity analyses for choice of coefficient priors (supplementary data “Sensitivity analysis of priors for the coefficients” and Figure S1), weakly informative priors were chosen and differed by model family. Priors for each of the models are shown in supplementary data (Table S3). Posterior distributions were drawn using four chains of 4000 to 10,000 iterations each (typically with 30% warm-up), depending on model complexity.

351 Counterfactual interventions

352 To simulate counterfactual interventions, we used posterior draws from the total effect
353 models to generate predicted probabilities of infection and predicted mean EPG under both
354 observed and intervened exposure conditions. For each intervention, we selected a target
355 exposure variable (travel frequency, water contact duration and activity type) and modified
356 it to reflect the desired intervention. Where an exposure had downstream effects on other
357 variables in the DAG (Figure 2A), these variables were also modified, for example if
358 someone doesn't travel to the lake, duration in the lake would also be set to zero.
359

360 The original model, fitted with the unmodified factual dataset using the appropriate
361 adjustment sets (Table S2) was then applied to both the factual and counterfactual datasets
362 to obtain posterior predictions. This preserved the natural variation in covariates needed to
363 estimate $P(Y|X)$ when applying the do-calculus, where Y denotes the outcome (binary
364 infection status or mean EPG) and X the exposure of interest.

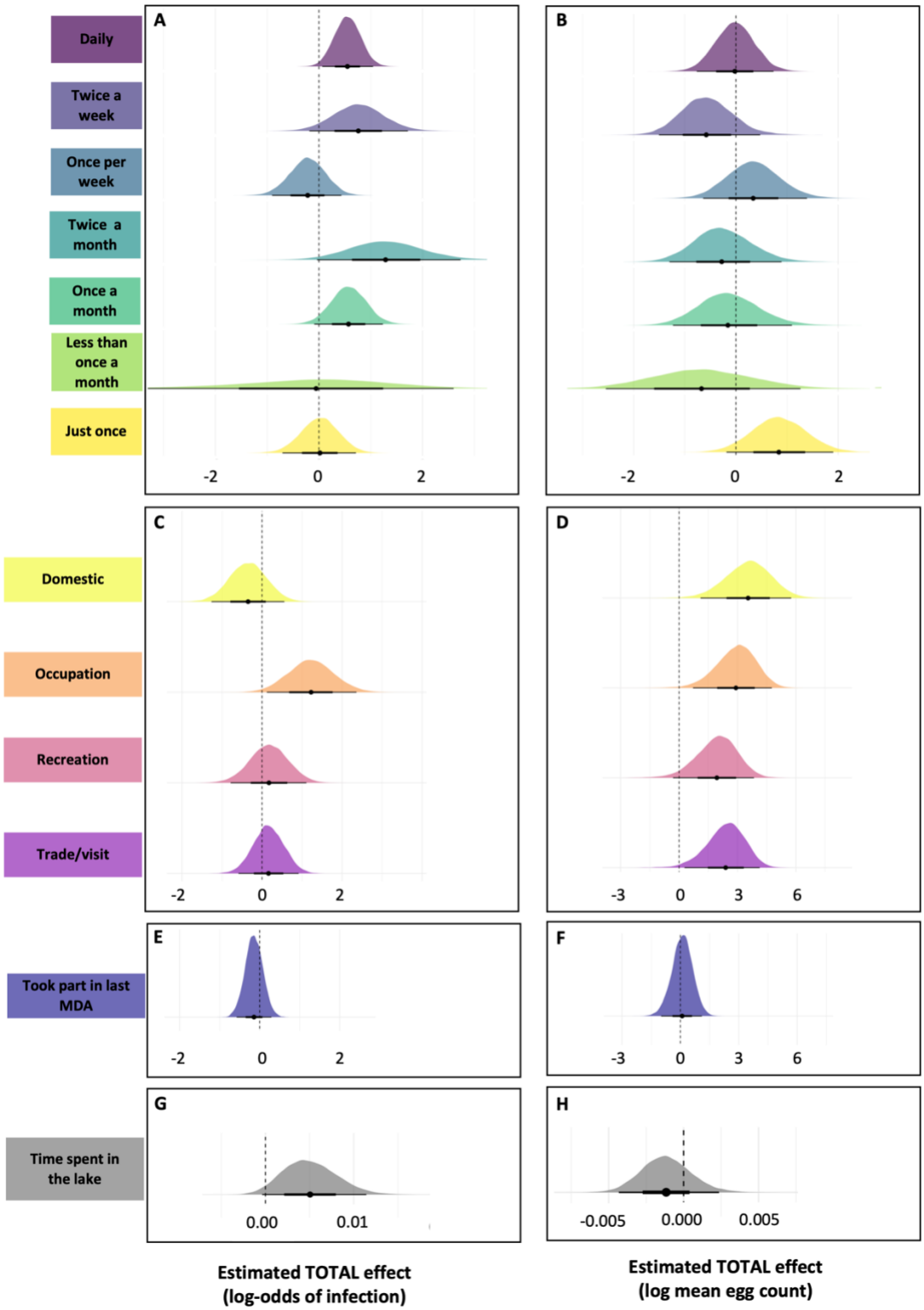
365 This approach enabled us to estimate how infection risk would change under hypothetical
366 interventions while avoiding post-treatment bias. We applied this approach to different
367 scenarios: A) **Travel frequency:** setting all daily travellers to no travel, examining (i) change
368 in infection probability for the entire population, and (ii) change only for daily travellers; B)
369 **Activity type:** (i) changing all "occupation: activity to "none" and examining infection
370 probability for the whole population and occupational travellers, (ii) changing all "domestic"
371 activity to "none" and examining mean EPG for the whole population and for domestic-
372 activity individuals only; and C) **Water contact duration:** setting duration to zero, examining
373 (i) change in infection probability for the entire population, and (ii) change only for those
374 who had reported lake contact.

375 For each scenario, posterior predictive distributions were summarised as means across
376 individuals for each posterior draw and visualised under observed and counterfactual
377 conditions using histograms. Presenting full distributions rather than single point estimates
378 allowed assessment of uncertainty, skew, and variability in predicted effects.

379 Mechanistic counterfactual analysis

380 To examine whether the effect of removing lake contact duration varied by travel
381 frequency, we conducted a mechanistic counterfactual analysis. Posterior infection risks
382 were predicted under the observed data and under a counterfactual scenario where
383 duration was set to zero for individuals who had reported >0mins in the Lake. Conditional
384 average treatment effects were then estimated by computing mean infection risks within
385 each travel-frequency group and taking the difference between counterfactual and
386 observed values. This approach quantified how the causal effect of removing lake contact
387 duration differed across levels of travel frequency, providing a sensitivity analysis of
388 heterogeneity in the duration-infection pathway.
389
390 **Results**
391
392 Prevalence of *Schistosoma mansoni* infections
393 Community-wide prevalence of infection was 51% in Nan, 45% in B, 33% Nam, 30% in Nw
394 and 27% in village S (mean=36%). No participant had an infection intensity of ≥400 EPG
395 (Table S4). As the WHO defines EPHP as <1% of SAC with ≥400 EPG, all villages had achieved
396 EPHP (2). For more detailed village, age, and sex level variation stratified by travel habits,
397 see supplementary Data (Table S4 and Figure S2).
398
399 Effect of travel frequency on infection
400 Among all participants, 56% (329/585) reported travelling to the lake at least once in the
401 last three months (Figure S3). Of these travellers, 41% (134/329) were infected and 30% of
402 those that did not travel to the lake were infected (105/256) (Table S5). Daily travel to the
403 lake was reported by 24% of participants (138/585). The mean EPG in each group (travel
404 and no travel) was 11.3 (sd = 22.2) and 11.6 (sd = 13.4) respectively.
405
406 The total effect of travel frequency on infection was estimated using the minimal
407 adjustment set of confounders only (Table S2). There was a positive total causal effect for
408 those who travelled once a month, twice a month, twice a week and daily to Lake Victoria in
409 the last three months, with daily visits having a significant 1.7x effect on increasing infection
410 risk (log odds= 0.55, 95% crI=0.06 to 1.05) (Figure 3A). When estimating the direct effect,

there was no longer a significant effect of daily visits (Figure S4), with the activities classed as occupation now having the only significant effect in the model (Figure 2A). Using a separate model of EPG, the amount of travel undertaken did not significantly affect the EPG in either the total or direct models (Figure 3B and Figure S5).



416

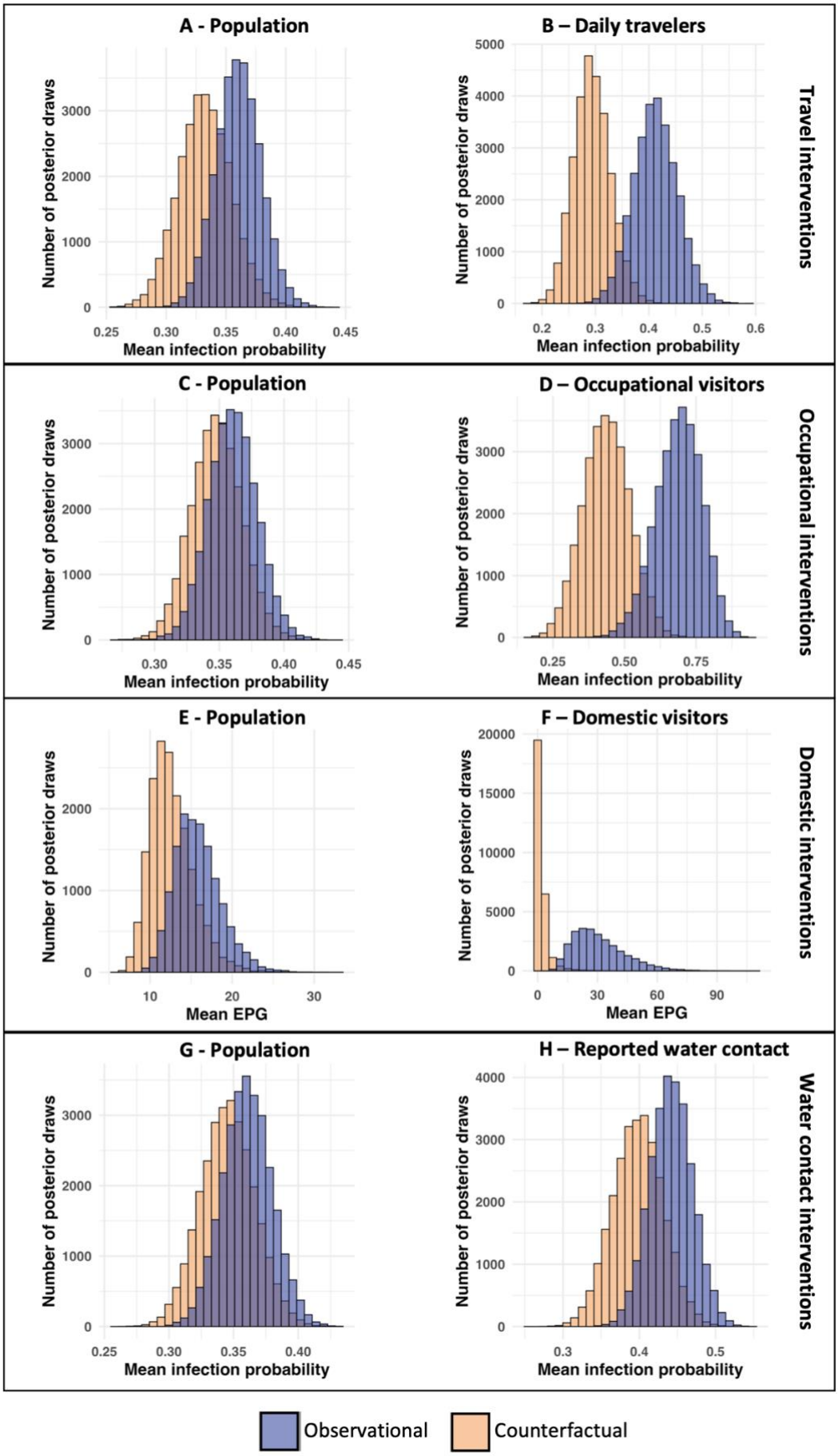
417 **Figure 3. Estimated total effects of lake-related behaviours on *Schistosoma mansoni***

418 **infection and infection burden.**

419 Each pair of panels (left = infection probability; right = infection burden) shows posterior
420 distributions from Bayesian regression models estimating the total effects of specific lake-
421 related variables.
422 Panels A & B: Self-reported travel frequency to Lake Victoria (relative to no travel) on (A) the
423 probability of *S. mansoni* infection and (B) infection burden measured as eggs per gram
424 (EPG) of stool).
425 Panels C & D: Self-reported lake-related activity type on (C) probability of infection and (D)
426 infection burden.
427 Panels E & F: Participation in the most recent mass drug administration (MDA) on (E)
428 infection probability and (F) infection burden.
429 Panels G & H: Time spent in contact with lake water on (G) infection probability and (H)
430 infection burden.
431 Densities represent marginal posterior distributions of log-odds (infection) or log-mean EPG
432 (burden). Thick and thin horizontal lines represent 66% and 95% credible intervals,
433 respectively. The vertical dashed line indicates the null effect (log-odds or log-mean = 0). In
434 panel A, travel frequency categories are ordered from highest (daily visits) to lowest (single
435 visit). Observed infection prevalence and intensity by behavioural and exposure variables
436 can be visualised on Figure S3.
437
438 Effect of activity carried out whilst visiting Lake Victoria
439 When estimating the total effect of activity on infection, individuals reporting occupational
440 water contact had higher log-odds of being infected compared to those with no lake contact
441 (log-odds = 1.23; 95% crI= 0.12 to 2.36). Other activities were not significant predictors of
442 infection risk (Figure 3C).
443
444 Including water contact duration in the model to estimate the direct effect reduced the
445 estimated effect of occupational activity on infection (log odds = 0.91; 95% crI= –0.26 to
446 2.10), compared to the model without duration (Figure S6). Duration itself was positively
447 associated with infection risk (log-odds = 0.01, crI=0.0 to 0.01), confirming partial mediation
448 as outlined in the DAG (Figure 2A).
449

450 All activities were associated with a higher mean EPG compared to those reporting no lake-
451 related activity. Domestic activities had the greatest effect with a 35x higher mean EPG than
452 those reporting no activity at the lake (log odds= 3.55, 95% crI= 1.10 to 5.76) (Figure 3D).
453 Adjusting for duration of lake exposure in a direct effect model, did not meaningfully alter
454 the association between activity type and EPG in comparison to the total effect model
455 (Figure S7), and duration was not a significant predictor.
456
457 Effect of taking part in the last MDA
458 Across all villages, 73% (425/585) of participants said they took part in the last MDA. Details
459 of responses to why a participant did not take part in the last MDA and what age groups did
460 take part can be found in supplementary data (Tables S6 And S7). When modelling the
461 effect of MDA participation on infection, the total and direct effects used the same
462 adjustment sets, as there were no recorded mediators (Table S2). Individuals who reported
463 participating in MDA had lower though non-significant log-odds of infection than those who
464 did not (log-odds= –0.14; 95% crI = -0.57 to 0.29) (Figure 3E). There was no association
465 between MDA and mean EPG (log-mean = 0.09; 95% crI= –0.99 to 1.11) (Figure 3F).
466
467 Effect of time spent in Lake Victoria
468 The longest lake visits were for occupational activities (mean time spent (minutes) = 92, sd =
469 161), recreation followed (mean= 33, sd = 32), then trade/visiting (mean= 18, sd = 34), with
470 domestic activities shortest (mean= 14, sd = 32). For daily visits, occupations were rare (n =
471 9) but longest (mean= 64, sd = 70), while domestic visits were most common (n = 73) and
472 shortest (mean= 5, sd = 12) (Table S8).
473
474 When modelling the effect of time spent in Lake Victoria on infection, the total and direct
475 effects used the same adjustment sets as there were no mediators (Table S2). The log-odds
476 of infection increased by 1% (log-odds = 0.01, 95% crI= -0.0 to 0.01) for every additional
477 minute (Figure 3G) (note that the lower bound is essentially on 0 suggesting an incalculably
478 small possibility of 0, and so technically only a weekly positive association). While this effect
479 is small per minute, it accumulates across longer durations, for example, an additional 60
480 minutes corresponds to approximately 82% higher odds of infection than someone who

481 spends no time in the lake (assuming linearity) (log-odds of (0.01* 0.6)). There was no effect
482 of duration on mean EPG (log-odd = -0.0, 95% crI= -0.0 to 0.0) (Figure 3H).
483
484 Counterfactual interventions
485 The first counterfactual simulation removed daily lake travel, reducing infection probability
486 by 3% in the whole population from 36% (95% crI: 32–40%) to 33% (95% crI: 29–38%)
487 (Figure 4A) and by 12% among daily travellers from 41% (95% crI: 33–50%) to 29% (95% crI:
488 23–36%) among daily travellers (Figure 4B).
489
490 Removing occupational lake visits reduced infection probability by 1% in the whole
491 population from 36% (95% CrI: 32–40%) to 35% (95% crI: 31–39%) (Figure 4C) by 26%
492 among those reporting occupational visits, from 70% (95% crI: 52–83%) to 44% (95% crI: 28–
493 59%).
494
495 Eliminating domestic water activities reduced mean infection intensity from 16 EPG (95%
496 crI: 11–22) to 12 EPG (95% crI: 9–19) in the overall population (Figure 4E) and from 29 EPG
497 (95% crI: 12–62) to 1 EPG (95% crI: 0.2–11) among those reporting domestic activities
498 (Figure 4F).
499
500 Setting lake contact duration to zero reduced infection probability from 36% (95% crI: 32–
501 40%) to 35% (95% crI: 31–39%) in the overall population (Figure 4G) and from 44% (95% crI:
502 39–50%) to 40% (95% crI: 33–46%) among those reporting contact (Figure 4H), a reduction
503 of 1% and 4%, respectively.
504


A - Population
B – Daily travelers
Travel interventions
C - Population
D – Occupational visitors
Occupational interventions
E - Population
F – Domestic visitors
Domestic interventions
G - Population
H – Reported water contact
Water contact interventions
Number of posterior draws
Mean infection probability
Mean EPG
Observational
Counterfactual

505

506

507 **Figure 4.** Counterfactual predictions of *Schistosoma mansoni* infection probability under
508 reduced travel and water contact scenarios. Posterior distributions of predicted infection
509 probabilities for selected subgroups under observed (blue) and counterfactual (orange)
510 conditions, based on a Bayesian logistic regression models.
511 **A and B)** Travel interventions: the predicted infection probability of no daily travel on the
512 entire population (A); and only daily travellers (B).
513 **C and D)** Occupational interventions: the predicted infection probability of no one doing
514 occupational activities on the entire population (C); and only those going for occupational
515 tasks (D).
516 **E and F)** Domestic interventions: the predicted mean eggs per gram of no one doing
517 domestic activities on the entire population (E); and only those going for domestic tasks (D).
518 **G and H)** Water contact interventions: the predicted infection probability of no water
519 contact on the entire population (G); and only those who reported making contact (H).
520 Distributions represent posterior samples of individual infection probabilities aggregated
521 across the sample population.
522

523 <u>Mechanistic counterfactual: contact duration by travel frequency</u>
524 To examine whether the effect of removing lake contact duration varied across levels of
525 travel frequency, we estimated conditional average treatment effects (CATEs). Reductions in
526 infection probability were minimal among non-travellers and infrequent visitors (≤ once a
527 month), highest amongst those travelling twice per month to twice per week (difference in
528 infection risk of six-10 percentage points – Table 1). Daily travellers, while most at risk
529 overall (Figure 3), showed comparatively small reductions (three percentage point reduction
530 – Table 1).
531

532 **Table 1.** Contact-only estimated change in infection probability under do(Duration = 0)
533

Travel frequency	**Median Δ (risk difference)**	**95% CrI (lower, upper)**
Non-travellers	−0.030 (≈ 3%)	−0.071, 0.002
Just Once	−0.029 (≈ 3%)	−0.065, 0.002

< Once per month	−0.034 (≈ 3%)	−0.079, 0.003
Once per month	−0.038 (≈ 4%)	−0.088, 0.003
Twice per month	−0.102 (≈ 10%)	−0.213, 0.011
Once per week	−0.058 (≈ 6%)	−0.133, 0.004
Twice per week	−0.082 (≈ 8%)	−0.158, 0.008
Daily	−0.031 (≈ 3%)	−0.071, 0.002

Discussion

This study aimed to test whether habitual movement of people living in low *S. mansoni* endemicity areas to high-transmission sites around Lake Victoria, Uganda, could be sustaining transmission in surrounding inland communities that have reached the WHO-defined threshold for EPHP (2). Our analysis revealed that frequent travel, particularly daily travel is a plausible and quantifiable pathway for sustaining *S. mansoni* transmission in low-prevalence settings. Critically, our results indicate that multiple overlapping exposures contribute to infection. It is therefore unlikely that transmission can be controlled by a single intervention, highlighting the need for a systems-thinking approach that recognises behavioural, environmental, and structural factors influencing transmission dynamics in EPHP settings.

To explore how behavioural interventions might influence infection risk and assess potential resurgence pathways**,** we simulated a series of counterfactual scenarios using the fully resolved SCM. The aim was not to advocate for behavioural restrictions, but to identify groups who might benefit most from enhanced surveillance or targeted treatment. Across all scenarios, predicted infection probabilities and mean EPG were lower than in the observed data. The largest reductions in infection were consistently observed within the specific subgroups targeted by each intervention, with relatively small effects at the population level. At the individual scale, such interventions could be considered successful, reducing infection probability by 12% in daily travellers, 26% in occupational visitors, and lowering mean EPG in those undertaking domestic water activities from 29 to 1. These

558 changes would not only improve individual health outcomes but also reduce the likelihood
559 of those individuals contributing to contamination of home-village water sources in the
560 future. However, the insubstantial impact at the population level highlights a limitation of
561 targeting individual risk factors to control transmission in established transmission settings.

562 Building on these modelled insights, the empirical data show that daily travel to Lake
563 Victoria was associated with a 1.7-fold higher probability of infection compared with no
564 travel, confirming a causal link between travel frequency and infection. When potential
565 mediators were included, this association became non-significant, indicating that the
566 elevated risk among daily travellers is largely explained by lake-related activities and time
567 spent in the water. Participants involved in occupational water contact had a 3.4-fold higher
568 infection probability than those reporting no lake activity, spending on average 92 minutes
569 in the lake compared with 14 minutes for domestic activities. Although domestic activities
570 were not linked to higher infection probability, they were strongly associated with higher
571 mean EPG among those infected, possibly reflecting occasional high-risk behaviours such as
572 washing clothes (22% of domestic cases) in shallow shoreline waters where Biomphalaria
573 snails and cercarial densities are highest. Weekly travellers showed no elevated infection
574 risk despite moderate exposure durations (~39–40 minutes), likely due to the low
575 proportion engaged in occupational contact (4%).
576
577 Mechanistic counterfactual analysis, simulating zero minutes of lake contact while
578 stratifying by travel frequency to assess the direct contribution of contact duration
579 independent of mobility behaviour, revealed the greatest reductions in infection probability
580 among those travelling twice per month, whereas daily travellers, the group at highest
581 overall risk showed smaller changes, likely reflecting residual infection in this group. Taken
582 together, these findings indicate that infection risk is shaped by the frequency, nature, and
583 duration of water contact, and that some transmission persists among individuals with little
584 or no lake contact. This further supports the earlier evidence that multiple overlapping
585 exposures contribute to infection risk. The detection of infections among participants who
586 never travelled to the lake supports the existence of inland transmission foci sustained by
587 parasite importation from mobile individuals. Frequent travellers therefore appear to play a
588 pivotal role in seeding infections within their home communities through routine water

contamination, allowing transmission to persist among non-travellers and across the wider area. This connectivity blurs the boundaries between “high-risk” and “low-risk” villages. Although all five study villages met EPHP criteria, with no heavy-intensity infections and low mean egg counts among egg-positive individuals (11 EPG, SD = 20), these low-intensity but persistent infections have important public health implications, highlighting the risk of underestimating transmission potential and resurgence in post-EPHP settings, suggesting that reliance on egg counts alone underestimates the true reservoir.

We therefore suggest that control in settings which have achieved EPHP require a systems-based approach that addresses more than just the proximal drivers of transmission (e.g. human behaviour and environmental reservoirs) to acknowledge the multiplicative effects of distal drivers that are often socio-political and structural in nature (35). Our statistical framework provides a tool to assess the potential impact of such strategies without relying on large, highly parameterised dynamic models, which require extensive, context-specific data. By linking empirical behavioural data to infection outcomes, this approach can help understand mechanistic pathways which could lead to resurgence. Looking forward, this framework could also inform adaptive post-EPHP surveillance and intervention strategies, enabling programmes to anticipate and respond dynamically to early signals of re-emergence.

<u>Limitations</u>

This study relied on self-reported behaviour from questionnaires, rather than direct observations. While direct observation would offer more accurate behavioural data, it was not feasible at the scale of this study. Misreporting is a known issue with self-reports (36), and one such example emerged in our data when participants were asked about water contact at or near their homes. Technicians working on this project had noted that many individuals used nearby swampland as a local water source, and so a set of questions was included to distinguish infection risk from local versus Lake Victoria exposure. However, upon examination of the reported home water sources, 50% were found to be on the shores of Lake Victoria, suggesting the question was misunderstood. As a result, these data, and responses to all home water contact questions, were excluded from the analysis due to concerns over validity.

621

622 Another limitation relates to interpretation of MDA participation. Although 74% of
623 participants reported taking part in the most recent MDA, the effect of treatment on
624 infection risk was weak, and no effect was observed on infection burden. This could be due
625 to rapid reinfection, or low drug efficacy, but may also reflect overreporting of participation.
626 This is a well-documented issue in qualitative research, where answers align with what the
627 participant believes is most desirable to the researcher (37). Furthermore, 50% of PSAC
628 reported taking part in last MDA, although were not eligible due to a lack of paediatric
629 praziquantel at the time. Attempts were made to retrieve the district reports on MDA, but it
630 was not possible, therefore the validity of self-reporting is without direct confirmation (e.g.
631 drug registers).

632

633 We hypothesised that high levels of habitual travel could also increase the risk of missing
634 MDA rounds, and our data supported this, with 28 individuals reporting they missed the last
635 MDA due to travel. However, if missing MDA is possible, so is missing surveys such as our
636 own study, and such individuals may be systematically under-represented in surveillance
637 and/or research activities, and may bias both epidemiological estimates and intervention
638 evaluations (38). In our setting, this implies that the infection risk and MDA non-
639 participation associated with high travel frequency may be underestimated, precisely
640 because the most mobile individuals may have been absent during data collection.

641

642 Conclusions

643 The findings presented here suggest that, in regions which have achieved EPHP, a small
644 subset of individuals engaging in high-risk behaviours may help to sustain local prevalence,
645 particularly occupational lake contact and habitual travel, and may serve as a bridge in
646 maintaining infection within otherwise low-risk groups. Counterfactual simulations showed
647 that while targeted interventions substantially reduce risk in individuals, their effect at the
648 population level is negligible, indicating that such approaches alone are insufficient.
649 Preventing resurgence will therefore require a systems-based strategy that integrates
650 human, environmental, structural, and ecological drivers of transmission to safeguard long-
651 term gains and achieve durable interruption of *S. mansoni* transmission.

652

Acknowledgements

We would like to thank the communities of, and individuals in, the study villages for participating in this study.

Data Availability

https://github.com/iamjessclark/travelStudySchisto

Competing Interests Statement

The authors declare no competing interests nor conflicts of interest.

Ethical Approval Statement

Study protocols were strictly adhered to following approval by the University of Glasgow MVLS ethics committee (200160068), Ministry of Health Uganda, Vector Control Division research ethics committee (VCDREC/062) and Uganda National Council for Science & Technology (UNCST-HS 2193).

Funding Statement

This work was supported by Wellcome [204820/Z/16/Z] awarded to JC, the European Research Council [SchistoPersist ERC starting grant 680088] awarded to PHLL and the Engineering and Physical Sciences Research Council [EP/T003618/1] awarded to PHLL. RML. supported by the Wellcome Trust (218492/Z/19/Z; supervised by ABP, JPW, and PHLL).

List of Figures and tables

Figure 1. Map of the study area showing the locations of the five participating villages along the northern shore of Lake Victoria, Uganda. Each point represents one village, colour-coded by name: B (purple), Nw (red), Nam (orange), Nan A (green), and S (pink). The left panel displays a detailed view of the study area, while the right panel provides a broader map of Uganda, with a black rectangle indicating the location of the zoomed-in map.

Figure 2 Directed acyclic graph (DAG) representing hypothesised causal relationships between variables and *Schistosoma mansoni* infection (A), with structural equations for

each node (B). (A) The DAG outlines the assumed causal structure linking demographic, behavioural, and environmental variables to infection. Arrows represent direct causal effects. Key exposure variables include travel frequency, activity type, and duration of water contact, with potential confounders such as age class, sex, and location. (B) Structural equations define the dependencies of each variable, with terms representing regression coefficients (β), unobserved residual error (U), and observed covariates. Ag = age class, S = sex, L = location, M = MDA, F = frequency of travel, A = activity, D = duration in the lake, I = infection.

Figure 3. Estimated total effects of lake-related behaviours on *Schistosoma mansoni* infection and infection burden.

Each pair of panels (left = infection risk; right = infection burden) shows posterior distributions from Bayesian regression models estimating the total effects of a specific lake-related variable.

Panels A & B: Self-reported travel frequency to Lake Victoria (relative to no travel) on (A) the probability of *S. mansoni* infection, and (B) infection burden (measured as eggs per gram (EPG) of stool).

Panels C & D: Self-reported lake-related activity type on (C) infection probability and (D) infection burden.

Panels E & F: Participation in the most recent mass drug administration (MDA) on (E) infection probability and (F) infection burden.

Panels G & H: Time spent in contact with lake water on (G) infection probability and (H) infection burden.

Densities represent marginal posterior distributions of log-odds (infection) or log-mean EPG (burden). Thick and thin horizontal lines represent 66% and 95% credible intervals, respectively. The vertical dashed line indicates the null effect (log-odds or log-mean = 0). In panel A, travel frequency categories are ordered from highest (daily visits) to lowest (single visit).

Figure 4. Counterfactual predictions of *Schistosoma mansoni* infection probability under reduced travel and water contact scenarios.

716 Posterior distributions of predicted infection probabilities for selected subgroups under
717 observed (blue) and counterfactual (orange) conditions, based on a Bayesian logistic
718 regression models.
719 **A and B)** Travel interventions: the predicted infection probability of no daily travel on the
720 entire population (A); and only daily travellers (B).
721 **C and D)** Occupational interventions: the predicted infection probability of no one doing
722 occupational activities on the entire population (C); and only those going for occupational
723 tasks (D).
724 **E and F)** Domestic interventions: the predicted mean eggs per gram of no one doing
725 domestic activities on the entire population (E); and only those going for domestic tasks (D).
726 **G and H)** Water contact interventions: the predicted infection probability of no water
727 contact on the entire population (G); and only those who reported making contact (H).
728 Distributions represent posterior samples of individual infection probabilities aggregated
729 across the sample population.
730
731 **Table 1.** Contact-only estimated change in infection probability under do(Duration = 0)
732
733
734 **<u>Supplementary Figures and Tables</u>**
735
736 **Table S1.** Participants responses to what activity they did at Lake Victoria put into groups for
737 statistical analysis
738
739 **Table S2.** Minimal adjustment sets used in total and direct effect models by exposure.
740
741 **Figure S1.** Total effects of activity on *Schistosoma mansoni* infection burden.
742 Posterior estimated total effects of lake-related activities, on the log mean of *S. mansoni*
743 eggs per gram of stool. The model included fixed effects for activity type (domestic,
744 occupational, recreational, trade/visit), and was adjusted for age class, sex, and location and
745 duration.
746 Posterior densities represent marginal distributions of estimated effects, with the vertical
747 dashed line indicating the null (log = 0).

A) Moderately informative coefficient priors of Normal(0,5)
B) Conservative coefficient priors of Normal(0,2)

Table S3. Model parameters and prior distributions.

Table S4. *Schistosoma mansoni* prevalence and mean eggs per gram (EPG) by travel frequency to Lake Victoria, Uganda

Table S5. Frequency of Lake Victoria travel and distribution of reported activities, stratified by age group, sex, and village.

Figure S2. Prevalence of *Schistosoma mansoni* by village and Sex. Prevalence measured by a positive Kato Katz from three duplicate slides or if a POC-CCA test had a G-score ≥3. F = female and M = male.

Figure S3. Observed infection prevalence and intensity by behavioural and exposure variables.

Observed infection prevalence (top and middle rows) and infection intensity (bottom row) plotted against behavioural and exposure variables included in the structural causal model. Bars and ribbons show observed mean prevalence with 95% binomial confidence intervals. Points represent individual infections (1 = infected, 0 = uninfected), shown with jitter for visibility.

(A) Infection prevalence by travel frequency to Lake Victoria (visits in the past three months).
(B) Infection prevalence by participation in the most recent mass drug administration (MDA).
(C) Infection prevalence by activity type at Lake Victoria.
(D) Infection prevalence by time spent in the lake (binned durations).
(E) Infection intensity (EPG, $\log_{1+}$ scale) by time spent in the lake (continuous).

(F–H) Infection intensity (EPG, $\log_{1+}$ scale) by travel frequency (F), MDA participation (G), and activity type (H).

Figure S4. Direct effect of travel frequency on log odds of infection with *Schistosoma mansoni.*

Posterior distributions for the estimated direct effects of self-reported travel frequency to Lake Victoria on the probability of infection, relative to individuals reporting no travel in the preceding three months (reference category). Estimates are log-odds coefficients from a Bayesian logistic regression model with a Bernoulli likelihood and logit link function. The model included fixed effects for travel frequency, age class, sex (binary), and location and the mediator's activity and MDA. Densities represent marginal posterior distributions for each travel frequency level, with the vertical dashed line indicating the null effect (log-odds = 0). Travel frequency levels are ordered from highest (92 days) to lowest (1 day) reported visits to the lake.

Figure S5. Direct effect of travel frequency on log mean on burden of *Schistosoma mansoni* eggs per gram in stool.

Posterior distributions for the estimated direct effects of self-reported travel frequency to Lake Victoria on mean eggs per gram of stool, relative to individuals reporting no travel in the preceding three months (reference category). Estimates are log mean coefficients from a Bayesian logistic regression model with a gamma likelihood and log link function. Densities represent marginal posterior distributions for each travel frequency level, with the vertical dashed line indicating the null effect (log-odds = 0). Travel frequency levels are ordered from highest (92 days) to lowest (1 day) reported visits to the lake.

Figure S6. Direct effects of activity on *Schistosoma mansoni* infection probability.

Posterior distributions for the estimated direct effects of lake-related activities, the log odds of infection. The model included fixed effects for activity type (domestic, occupational, recreational, trade/visit), and was adjusted for age class, sex, and location and duration. Posterior densities represent marginal distributions of estimated effects, with the vertical dashed line indicating the null (log = 0).

808

809 **Figure S7.** Direct effects of activity on *Schistosoma mansoni* infection burden.

810 Posterior distributions for the estimated direct effects of lake-related activities, on the log
811 mean of *S. mansoni* eggs per gram of stool. The model included fixed effects for activity type
812 (domestic, occupational, recreational, trade/visit), and was adjusted for age class, sex, and
813 location and duration.

814 Posterior densities represent marginal distributions of estimated effects, with the vertical
815 dashed line indicating the null (log = 0).

816

817 **Table S6.** Self-reported reasons for not taking part in the last mass drug administration.

818

819 **Table S7.** Self-reported participation in the last MDA by age group.

820

821 **Table S8.** Frequency of Lake Victoria travel and distribution of reported activities, stratified
822 by duration in the lake.

823 **References**

824

825 1. Meurs L, Mbow M, Boon N, van den Broeck F, Vereecken K, Dièye TN, et al. Micro-
826 geographical heterogeneity in Schistosoma mansoni and S. haematobium infection and
827 morbidity in a co-endemic community in Northern Senegal. PLoS Negl Trop Dis. 2013 Dec
828 26;7(12):e2608.

829 2. World Health Organization. WHO guideline on control and elimination of human
830 schistosomiasis. 2022.

831 3. Webster JP, Molyneux DH, Hotez PJ, Fenwick A. The contribution of mass drug
832 administration to global health: past, present and future. Philos Trans R Soc Lond B Biol
833 Sci. 2014;369(1645):20130434.

834 4. Casulli A. New global targets for NTDs in the WHO roadmap 2021–2030. 2021;

835 5. Katz N, Chaves A, Pellegrino J. A simple device for quantitative stool thick-smear
836 technique in Schistosomiasis mansoni. Rev Inst Med Trop São Paulo. 1972;14(6):397–
837 400.

838 6. World Health Organization,. The control of schistosomiasis. Second report of the WHO
839 Expert Committee. 1993 p. 1–86. Report No.: WHO TRS N°830.

840 7. Ayabina DV, Clark J, Bayley H, Lamberton PHL, Toor J, Hollingsworth TD. Gender-related
841 differences in prevalence, intensity and associated risk factors of Schistosoma infections
842 in Africa: A systematic review and meta-analysis. PLoS Negl Trop Dis. 2021 Nov
843 17;15(11):e0009083.

844 8. Toor J, Truscott JE, Werkman M, Turner HC, Phillips AE, King CH, et al. Determining post-
845 treatment surveillance criteria for predicting the elimination of Schistosoma mansoni
846 transmission. Parasit Vectors. 2019 Sept 16;12(1):437.

847 9. Wang W, Bergquist R, King CH, Yang K. Elimination of schistosomiasis in China: Current
848 status and future prospects. PLoS Negl Trop Dis. 2021 Aug 5;15(8):e0009578.

10. Trippler L, Ame SM, Hattendorf J, Juma S, Abubakar S, Ali SM, et al. Impact of seven years of mass drug administration and recrudescence of Schistosoma haematobium infections after one year of treatment gap in Zanzibar: Repeated cross-sectional studies. PLoS Negl Trop Dis. 2021 Feb;15(2).

11. Liang S, Yang C, Zhong B, Qiu D. Re-emerging schistosomiasis in hilly and mountainous areas of Sichuan, China. Bull World Health Organ. 2006 Feb;84(2):139–44.

12. Arinaitwe M, Adriko M, Kibwika B, Tukahebwa EM, Faust CL, Lamberton PHL. Residence time, water contact, and age-driven Schistosoma mansoni infection in hotspot communities in Uganda. Am J Trop Med Hyg. 2021 Dec;105(6):1772–81.

13. Clark J, Arinaitwe M, Nankasi A, Faust CL, Moses A, Ajambo D, et al. Reconciling Egg- and Antigen-Based Estimates of Schistosoma mansoni Clearance and Reinfection: A Modeling Study. Clin Infect Dis Off Publ Infect Dis Soc Am. 2021 Aug 6;74(9):1557–63.

14. Exum NG, Kibira SPS, Ssenyonga R, Nobili J, Shannon AK, Ssempebwa JC, 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. 2019 Aug 14;13(8):e0007617.

15. Grobusch MP, Mühlberger N, Jelinek T, Bisoffi Z, Corachán M, Harms G, et al. Imported Schistosomiasis in Europe: Sentinel Surveillance Data from TropNetEurop. J Travel Med. 2003 June 1;10(3):164–9.

16. Cui J, Jiang ,Peng, Song ,Yan Yan, Zhang ,Xi, and Wang ZQ. Imported African schistosomiasis and the potential risk of transmission in China. Pathog Glob Health. 2018 Feb 17;112(2):101–5.

17. Steverding D. The spreading of parasites by human migratory activities. Virulence. 2020 Dec 31;11(1):1177–91.

18. Wang L, Wu X, Li X, Zheng X, Wang F, Qi Z, et al. Imported Schistosomiasis: A New Public Health Challenge for China. Front Med [Internet]. 2020 Oct 22 [cited 2025 May 27];7.

875 Available from:
876 https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2020.553487/full

877 19. Mari L, Gatto M, Ciddio M, Dia ED, Sokolow SH, De Leo GA, et al. Big-data-driven
878 modeling unveils country-wide drivers of endemic schistosomiasis. Sci Rep. 2017 Mar
879 28;7(1):489.

880 20. Perez-Saez J, Mari L, Bertuzzo E, Casagrandi R, Sokolow SH, Leo GAD, et al. A Theoretical
881 Analysis of the Geography of Schistosomiasis in Burkina Faso Highlights the Roles of
882 Human Mobility and Water Resources Development in Disease Transmission. PLoS Negl
883 Trop Dis. 2015 Oct 29;9(10):e0004127.

884 21. Ciddio M, Mari L, Sokolow SH, De Leo GA, Casagrandi R, Gatto M. The spatial spread of
885 schistosomiasis: A multidimensional network model applied to Saint-Louis region,
886 Senegal. Adv Water Resour. 2017 Oct 1;108:406–15.

887 22. Reitzug F, Ledien J, Chami GF. Associations of water contact frequency, duration, and
888 activities with schistosome infection risk: A systematic review and meta-analysis. PLoS
889 Negl Trop Dis. 2023 June 14;17(6):e0011377.

890 23. Hernan M. A definition of causal effect for epidemiological research. J Epidemiol
891 Community Health. 2004 Apr;58(4):265–71.

892 24. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in
893 epidemiology: the need for a pluralistic approach. Int J Epidemiol. 2016 Dec;45(6):1776–
894 86.

895 25. Casacuberta-Partal M, Hoekstra PT, Kornelis D, van Lieshout L, van Dam GJ. An innovative
896 and user-friendly scoring system for standardised quantitative interpretation of the
897 urine-based point-of-care strip test (POC-CCA) for the diagnosis of intestinal
898 schistosomiasis: a proof-of-concept study. Acta Trop. 2019 Nov;199:105150.

899 26. Colley DG, Binder S, Campbell C, King CH, Tchuenté LAT, N'Goran EK, et al. A Five-Country
900 Evaluation of a Point-of-Care Circulating Cathodic Antigen Urine Assay for the Prevalence
901 of Schistosoma mansoni. Am J Trop Med Hyg. 2013 Mar 6;88(3):426–32.

902 27. Clark J, Arinaitwe M, Nankasi A, Faust CL, Adriko M, Ajambo D, et al. Translating from egg-
903 to antigen-based indicators for Schistosoma mansoni elimination targets: A Bayesian
904 latent class analysis study. Front Trop Dis. 2022;3.

905 28. Corstjens P, Hoekstra PT, de Dood CJ, van Dam GJ. Utilizing the ultrasensitive
906 Schistosoma up-converting phosphor lateral flow circulating anodic antigen (UCP-LF CAA)
907 assay for sample pooling-strategies. Infect Dis POVERTY. 2017 Nov 1;6.

908 29. Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence
909 synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic
910 method for building directed acyclic graphs. Int J Epidemiol. 2020 Feb 1;49(1):322–9.

911 30. Textor J, Van Der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH. Robust causal
912 inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol. 2017 Jan
913 15;dyw341.

914 31. Petersen ML, van der Laan MJ. Causal Models and Learning from Data. Epidemiol Camb
915 Mass. 2014 May;25(3):418–26.

916 32. Pearl J, Glymour M, Jewell NP. Causal Inference in Statistics: A Primer. Wiley; 2016.

917 33. Pearl J. Causal inference in statistics: An overview [Internet]. Vol. 3. 2009 [cited 2025 Apr
918 27]. Available from: https://projecteuclid.org/journals/statistics-surveys/volume-3/issue-
919 none/Causal-inference-in-statistics-An-overview/10.1214/09-SS057.full

920 34. Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. J Stat Softw.
921 2017 Aug 29;80:1–28.

922 35. Makau-Barasa LK, Kamara K, Karuta C, Aderogba M, Leaning E, Bockarie M. Systems
923 thinking in the prevention, control, and elimination of neglected tropical diseases (NTDs).
924 Int J Infect Dis. 2025 Mar;152.

925 36. Lira B, O'Brien JM, Peña PA, Galla BM, D'Mello S, Yeager DS, et al. Large studies reveal
926 how reference bias limits policy applications of self-report measures. Sci Rep. 2022 Nov
927 10;12(1):19189.

928 37. Bergen N, Labonté R. “Everything Is Perfect, and We Have No Problems”: Detecting and
929 Limiting Social Desirability Bias in Qualitative Research. Qual Health Res. 2020 Apr
930 1;30(5):783–92.

931 38. Clark J, Davis EL, Prada JM, Gass K, Krentel A, Hollingsworth TD. How correlations
932 between treatment access and surveillance inclusion impact neglected tropical disease
933 monitoring and evaluation—A simulated study. PLoS Negl Trop Dis. 2023 Sept
934 6;17(9):e0011582.

935

936

937