

The Epidemiology of *Schistosoma mansoni*: Analysing the Influence of Current Water Contact, Safe Water Supplies, and Human Mobility in Rural Uganda



Fabian Reitzug

Balliol College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary Term, 2025

Acknowledgements

I would like to thank my parents, Claudia and Axel, for their unwavering support, love, and encouragement throughout my educational journey. To Sophia, Hannah, and Stefan, for their healthy perspectives and experience, which helped me keep my cool in moments of stress.

To Trishna, who, more than anyone else, made my time in Oxford truly wonderful and special.

I'm incredibly grateful to my office mates, especially Ellie and Sandrena, who filled the office with good vibes, laughter, and silliness.

Thanks to my Oxford friends—Bienfait, Chang, Josefina, Liliana, Liam, Mathis, Quentin, Sofia, Tin—with whom I shared so many wonderful memories. To my friends from afar—Bjørn, Caroline, Clements, Hannah, Juliette, Laure, Niels, Tom—with whom I shared the joys of Oxford when they came to visit.

To my supervisors, Golette and Renaud, for their guidance and mentorship, as well as my examiners, Deirdre and Michael, for their thoughtful feedback and constructive criticism.

To the Nuffield Department of Population Health for generous funding and support throughout.

And finally, a special thanks to Max and the Ugandan team who were there through all the good and tough moments in the field.

Abstract

Background: Schistosomiasis is a waterborne parasitic disease that affects 250 million people globally, more than 90% of whom live in sub-Saharan Africa. Human behaviour is a key influencing factor of local transmission dynamics yet remains poorly understood, with existing research being reliant on aggregate mobility data and self-reported water contact without validated measurement tools to study exposure behaviour. This thesis uses detailed individual-level survey, GPS logger, and direct observation data to characterise the influence of current human water contact behaviour, mobility, and access to safe water supplies on the epidemiology of *Schistosoma mansoni* with a focus on rural Uganda. Here, current water contact is defined as engaging in any water contact within the past year or more frequently.

Methods: Chapter 2 presents a systematic review and meta-analysis of 101 studies covering 192,691 participants, which examines the relationship between current water contact and schistosome infection likelihood. Chapter 3 presents an analysis of water contact patterns in the SchistoTrack cohort (n = 2867) in Uganda, employing Bayesian variable selection and multivariable logistic regression. Chapter 4 utilises GPS loggers from 452 SchistoTrack participants to build spatial decay models for predicting water site usage patterns, with a focus on the influence of human mobility and safe water supplies.

Results: Chapter 2 found that individuals with current water contact had 3.14 times higher odds of schistosome infection compared to those without contact, and that associations varied based on climate. Chapter 3 identified older age, fishing occupations, proximity to multiple water sites, and village-level infection prevalence as key determinants of water contact. There was a strong age and gender dependence in water contact with minimal correlation between current water contact and infection. Chapter 4 developed spatial decay models that predict water site and tap usage with high accuracy and suggested that the presence of safe water infrastructure (taps/boreholes) could reduce water site usage by up to 10% and the duration of open water contact by up to 31%.

Significance: This thesis offers insights into the role of human behaviour for schistosomiasis; it challenges the assumption that current water contact is necessarily a strong proxy indicator for infection status. It highlights the strong spatial concentration of water contact among individuals living near water bodies and exponential distance decays in water contact. Generalisable models are provided that predict water site usage to aid future intervention targeting and evaluation for the provision of safe water supplies. These findings are anticipated to inform the improved integration of behavioural factors into transmission models and guide future exposure measurement in control programmes.

Contents

1	Introduction	1
1.1	Context and Global Burden of Schistosomiasis	1
1.2	Schistosome Biology	2
1.2.1	Immunology	3
1.3	Behavioural Aspects of Transmission	3
1.3.1	Human Water Contact Behaviour	4
1.3.2	Human Behaviour and Water Site Contamination	5
1.3.3	Human Mobility	5
1.4	Relevance of Human Behavioural Factors for Schistosomiasis Control Interventions and Policy	6
1.5	Relevance of Behavioural Factors for Mathematical Transmission Models	7
1.6	Research Questions	8
1.7	Thesis Aims and Content	9
1.8	Publications Arising from this Thesis	10
1.9	Contributions and Funding	10
1.10	Ethics Approval	11
2	Associations of Water Contact with Schistosome Infection Risk: A Systematic Review and Meta-Analysis	17
2.1	Background	17
2.2	Methods	19
2.2.1	Protocol and Search Strategy	19
2.2.2	Exposures and Outcome	20
2.2.3	Inclusion and Exclusion Criteria	20
2.2.4	Screening and Data Extraction	21
2.2.5	Subgroup Variables	21
2.2.6	Data Analysis	22
2.2.7	Risk of Bias	22
2.3	Results	24
2.3.1	Study Characteristics	25
2.3.2	Association of Having Any Water Contact with Infection Status	28
2.3.3	Association of Water Contact Frequency and Duration with Infection Status	32
2.3.4	Association of Water Contact Activities with Infection Status	33
2.4	Discussion	35
2.4.1	Any Water Contact	35
2.4.2	Water Contact Frequency and Duration	38
2.4.3	Water Contact Activities	39
2.4.4	WASH Thresholds and Community Infection Prevalence	39
2.4.5	Limitations	42
2.4.6	Conclusion	43
2.5	Supplementary Information	49

3	A Data-Driven Analysis of the Determinants of Current Water Contact and <i>S. mansoni</i> Infection in the SchistoTrack Study	63
3.1	Background	63
3.2	Methods	65
3.2.1	Study Setting, Participant Sampling, and Assessments	65
3.2.2	Fieldwork During the DPhil	66
3.2.3	Water Contact Outcome	68
3.2.4	<i>S. mansoni</i> Infection Outcome	69
3.2.5	Individual, Household, and Village Human–Environmental Variables	69
3.2.6	Comparison of Self-Reported Water Contact with Direct Water Contact Observations	71
3.2.7	Geospatial Data	72
3.2.8	Statistical Analysis	72
3.3	Results	75
3.3.1	Water Contact Type, Timing, Frequency, and Duration	75
3.3.2	Water Contact Dependence on Distance to Water Sites	76
3.3.3	Gender and Age-Specific Water Contact Patterns	78
3.3.4	Relevant Human-Environmental Dimensions of Water Contact	82
3.3.5	Determinants of Any Water Contact	82
3.3.6	Determinants of Water Contact Versus Infection	84
3.3.7	Validation of Water Contact and Infection Measurements	86
3.3.8	Predictive Capacity of Water Contact Models by Variable Selection Technique	87
3.4	Discussion	88
3.4.1	Limitations	92
3.4.2	Conclusion	92
3.5	Supplementary Information	99
4	The Influence of Safe Water Supplies and Human Mobility on Water Contact Patterns in Rural Uganda: A Fine-Scale Spatial Modelling Study with Wearable GPS Loggers	127
4.1	Background	127
4.2	Methods	131
4.2.1	Study Setting and Participant Sampling	131
4.2.2	GPS Loggers and Participant Flow	131
4.2.3	GPS Logger Fieldwork and Training During the DPhil	132
4.2.4	GPS Logger Testing and Validation	133
4.2.5	GPS Logger Data Retrieval	133
4.2.6	Water Site and Tap/Borehole Data	134
4.2.7	Deriving Water Contacts and Tap/Borehole Visits	134
4.2.8	Sanitation Infrastructure	135
4.2.9	Dataset and Variables	135
4.2.10	Human Mobility	137
4.2.11	General Spatial Decay Modelling Framework	137
4.2.12	Spatial Decay Models of Water Site and Tap/Borehole Usage	139
4.2.12.1	Any usage models	139
4.2.12.2	Usage duration models	140
4.2.13	Statistical Analysis Methods	141
4.2.14	Out-of-Sample Prediction	143
4.2.15	Sensitivity and Robustness Analyses	144
4.3	Results	145
4.3.1	Study Overview	145
4.3.2	Water Site Usage	146

4.3.3	Tap/Borehole Usage	147
4.3.4	Combined Usage of Water Sites and Taps/Boreholes	148
4.3.5	Demographic Differences in Water Site and Tap/Borehole Usage	148
4.3.6	Water Site and Tap/Borehole Usage Over Household Distance	149
4.3.7	Water Site and Tap/Borehole and Human Mobility	149
4.3.8	Logistic Regression Models for Predicting Water Site Usage	154
4.3.9	Spatial Decay Models for Predicting Water Site and Tap/Borehole Usage	155
4.3.10	Spatial Decay Models for Predicting Water Site and Tap/Borehole Usage Duration	157
4.3.11	Spatial Decay Models Accounting for the Number of Available Water Sites	158
4.3.12	Crowding Out of Water Site Usage by Tap/Borehole Usage	159
4.3.13	Crowding Out of Water Site Usage Duration by Tap/Borehole Usage	161
4.3.14	Population-Wide Predictions of Water Site Usage	163
4.3.15	Minimum Data Needs	164
4.3.16	Robustness to Sampling Frequency	164
4.3.17	Representativeness	165
4.3.18	Wear Non-Compliance	166
4.4	Discussion	167
4.4.1	Limitations	170
4.4.2	Conclusion	172
4.5	Supplementary Information	177
5	Conclusion	185
5.1	Thesis Aims and Contributions	185
5.1.1	Aims	185
5.1.2	Main Contributions	185
5.2	Summary of Main Findings	186
5.2.1	The Complex Relationship Between Water Contact and Schistosome Infection	186
5.2.1.1	Associations between current water contact and infection were weak	186
5.2.1.2	Water contact and schistosomiasis hotspots	187
5.2.2	Key Determinants of Water Contact	188
5.2.2.1	Age-specific water contact patterns	188
5.2.2.2	Gender-specific water contact patterns	188
5.2.2.3	Impact of human mobility on water contact	189
5.2.3	Environmental and Spatial Factors	189
5.2.3.1	Distance to waterbodies	189
5.2.3.2	Water site characteristics and contamination	190
5.2.4	WASH Factors	190
5.2.4.1	Safe sanitation infrastructure	190
5.2.4.2	Safe water supplies	191
5.3	Implications for Water Contact Measurement, Modelling, and study design	191
5.3.1	Water Contact Measurement	191
5.3.1.1	Any water contact	192
5.3.1.2	Water contact frequency and duration	192
5.3.2	Modelling Water Contact	193
5.3.2.1	Predictability and determinants of water contact	193
5.3.2.2	Using water site distance for predicting water contact	194
5.4	Implications for Control Strategies	194
5.4.1	MDA	194
5.4.2	Exposure Reduction Interventions	195
5.4.3	Environmental Control	195
5.4.4	Safe Water Supplies	196

5.5	Generalisability	196
5.6	Thesis Limitations	197
5.7	Conclusion and Future Directions	199

Bibliography	204
---------------------	------------

List of Figures

1.1	The <i>Schistosoma</i> sp. lifecycle, human behaviour, and control strategies.	4
2.1	PRISMA Flowchart	24
2.2	Distribution of studies across geography and time	27
2.3	Meta-analysis of the association between individual-level water contact and odds of schistosome infection	28
2.4	Funnel plot for meta-analysis of the association between water contact and schistosome infection	29
2.5	Subgroup meta-analysis of the association between water contact and schistosome infection	31
2.6	Meta-analysis of the association between daily, weekly, and monthly water contact and schistosome infection	32
2.7	Meta-analysis of the association between individual-level water contact duration and odds of schistosome infection	33
2.8	Meta-analysis of the associations between water contact activities and odds of schistosome infection	34
S2.1	Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis of adjusted versus unadjusted effect sizes.	49
S2.2	Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis of infection/reinfection.	50
S2.3	Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis by study quality (low, moderate, high).	51
S2.4	Power analysis for meta-analysis of the association between individual-level water contact and odds of schistosome infection.	52
S2.5	Influence analysis of any single study in the meta-analysis of the association between individual-level water contact and odds of schistosome infection.	53
S2.6	Subgroup meta-analysis by study setting of the association between individual-level water contact and odds of schistosome infection in children.	54
S2.7	Subgroup meta-analysis of the association between individual-level water contact and odds of schistosome infection in community-based studies.	55
S2.8	Funnel plots for meta-analysis of the association between daily, weekly, monthly water contact and odds of schistosome infection.	56
S2.9	Influence analysis of any single study in the meta-analysis of the association between daily, weekly, monthly water contact and odds of schistosome infection.	57
S2.10	Funnel plot for meta-analysis of the association between individual-level water contact duration and odds of schistosome infection	58
S2.11	Funnel plot for meta-analysis of the associations between water contact activities and odds of schistosome infection	59

S2.12	Influence analysis of any single study in the meta-analysis of the associations between water contact activities and odds of schistosome infection (1/2)	60
S2.13	Influence analysis of any single study in the meta-analysis of the associations between water contact activities and odds of schistosome infection (2/2)	61
3.1	SchistoTrack study overview	75
3.2	SchistoTrack study areas	77
3.3	Proportion of individuals with water contact over household waterbody distance	78
3.4	Variation in water contact and infection over age	79
3.5	Composition of water contact activities over age and gender	80
3.6	Relevant dimensions of water contact and infection	81
3.7	Determinants of water contact	83
3.8	Determinants of infection status and heavy infection intensity	86
3.9	Predictive capacity of water contact and infection models	87
S3.1	Distribution of water contact frequency and duration	111
S3.2	Distribution of water contact and participants over distance to water sites	112
S3.3	Variation in water contact over gender and age	113
S3.4	Model of water contact with occupation or gender removed	114
S3.5	Models of occupational, domestic, and recreational water contact	115
S3.6	Model of water contact frequency	116
S3.7	Model of water contact duration	117
S3.8	Variation in infection by water contact and age	118
S3.9	Variation in water contact and infection over waterbody distance	119
S3.10	Model of infection intensity	120
S3.11	Comparison of self-reported and direct observation data	121
S3.12	Sensitivity to infection outcome used	122
S3.13	Participant flow diagram	123
S3.14	Comparison of BAS and LRT results by variable types and levels	124
S3.15	Model of infection with additional snail variables	125
S3.16	Model of infection with additional water contact variables	126
4.1	Participant flowchart	146
4.2	Schematic of deriving water site and tap/borehole usage from GPS logger data	147
4.3	Water site and tap/borehole visits over age and distance	150
4.4	Spatial decay models for water site and tap/borehole usage	156
4.5	Spatial decay models for water site and tap/borehole usage duration	157
4.6	Spatial Decay Models Accounting for Site Order	159
4.7	Comparison of spatial decay models with and without crowding for predicting water site usage	160
4.8	Crowding out effects of taps/boreholes on water site usage duration	162
4.9	Population-wide predictions of water site usage	164
4.10	Minimum data needs for estimating the spatial decay	165
S4.1	GPS logger in waterproof pouch and with attached lanyard.	179
S4.2	GPS loggers with water damage after fieldwork.	179
S4.3	Comparison between power law decay, exponential decay, and a Hill function in approximating the global distance decay in water site usage. Black dots represent the true proportion of individuals using water sites, calculated in 100m increments.	180
S4.4	Proportion of water site contact and tap/borehole usage episodes by time of day and district.	181
S4.5	Density plot of the district-specific mobility terciles.	182
S4.6	GAMs of distance to frequented water sites and taps/boreholes over age and mobility tercile	182

S4.7	Logistic regression model versus spatial decay model	183
S4.8	Water site usage models incorporating district and mobility	183
S4.9	Tap/borehole usage models incorporating district	184
S4.10	The impact of removing short-duration usage events.	184

List of Tables

2.1	Study characteristics summary	25
S3.1	Variable definitions	99
S3.2	Key participant characteristics	106
S3.3	Water contact per activity	107
S3.4	Water contact frequency per activity (in trips per week)	108
S3.5	Water contact duration per activity (in hrs per week)	108
S3.6	Water contact at high-risk time per activity	109
S3.7	Comparison of variable selection between LRTs and BVS	110
4.1	Summary of participant characteristics	151
4.2	Summary of participant characteristics	153
4.3	Performance of logistic regression models (baseline models)	154
4.4	Model selection for spatial decay models	155
4.5	Comparison of participants with sufficient GPS data versus those with missing data and the wider SchistoTrack study population.	166
S4.1	Number of the GPS points per participant per day, by district.	177
S4.2	Comparison of participants by number of recorded GPS points, grouped into terciles.	178

Chapter 1

Introduction

1.1 Context and Global Burden of Schistosomiasis

Human schistosomiasis is a neglected tropical disease that is endemic to 78 countries across Asia, the Middle East, North Africa, and South America [1]. According to the 2021 Global Burden of Disease Study, 1,746,000 disability-adjusted life-years (DALYs) and 13,000 annual deaths are attributed to schistosomiasis [2, 3]. Estimates of the number of people infected vary between 150–250 million people [1, 3, 4]. Almost 90% of all cases, morbidity, and mortality are estimated to occur in sub-Saharan Africa [2, 3] where 288 million people were estimated to be at risk of schistosomiasis in 2019 and where 128–163 million people were infected, according to recent modelling studies [3, 5]. Schistosomiasis is a disease of poverty, with rural populations commonly experiencing the highest levels of exposure due to subsistence fishing or agriculture and the lack of access to safe water, sanitation, and hygiene (WASH) infrastructure [6].

For children, untreated childhood infections have been associated with malnutrition, stunting, and delayed cognitive development [7–9]. For adults, chronic inflammation due to sustained infection has been shown to cause liver, spleen, and bladder damage and to increase the risk of cancers or organ failure (depending on the *Schistosoma* species) [10]. For communities, the absence of treatment has been linked to reduced fishing and agricultural productivity [11, 12]. Schistosomiasis can contribute to poverty traps by reinforcing cycles of disease and economic hardship [13]. For example, children in poor households may be more likely to be malnourished and chronically infected and thus at increased risk of developing anaemia, stunting, and delayed cognitive development, which can hinder their ability to attend and perform well in school [14]. Reduced educational attainment limits future employment opportunities and earning potential,

increasing the likelihood of persistent intergenerational poverty traps.

1.2 Schistosome Biology

The causative agents of schistosomiasis are parasitic flatworms (trematodes) of the genus *Schistosoma*. In sub-Saharan Africa, the geographic focus of this thesis, the two main species are *S. haematobium* and *S. mansoni* [15, 16]. *S. haematobium* causes the urogenital form of schistosomiasis, while *S. mansoni* causes the intestinal form. These species exhibit slightly different life cycles with different competent snail hosts required, different organs targeted (liver, spleen, and gut for intestinal species whereas predominantly the bladder for the urogenital forms), and different egg excretion routes [1].

When an infected individual contaminates the environment via open urination (*S. haematobium*) or open defecation (*S. mansoni*), the parasite eggs are released into the environment. Upon contact with freshwater, the eggs hatch miracidia, the first larval stage of the parasite. The miracidia penetrate freshwater snails, where they undergo asexual reproduction. After 30 days of development within the intermediate hosts [17], the snails shed cercariae, the second larval stage, in response to sunlight exposure. The cercariae, which are free-swimming, locate human hosts via thermosensation and penetrate the skin. During this process, the cercariae lose their tails and transform into schistosomulae.

Once inside the host, the schistosomulae enter the venous blood system and migrate to the hepatic portal system. Inside the mesenteric veins of the liver, the schistosomulae mature into sex-differentiated adult flukes. Male-female fluke pairs are required for reproduction. Once paired, the flukes migrate; for the species *S. mansoni* to the mesenteric veins of the bowel, where they attach to the intestinal wall. *S. haematobium* flukes attach to the venous plexus of the bladder and urogenital tract. Four to six weeks post-infection, the eggs laid by the female penetrate through the intestinal wall (*S. mansoni*) or bladder wall (*S. haematobium*) and are excreted via the faeces or urine, respectively, completing the lifecycle [1, 18].

1.2.1 Immunology

Infection prevalence with schistosomiasis typically peaks in children aged 10–15 years and declines thereafter in endemic areas. Most individuals become infected early in life, as demonstrated by a study in Kenya where seroprevalence based on antibody responses to *S. mansoni* soluble egg antigen reached 38% by age two and 63% by age five [19].

This natural decline in infection levels with age has been attributed to the gradual development of acquired immunity [20]. Repeated exposure to infection and treatment with praziquantel (which kills adult worms and releases their antigens) both contribute to protective immune responses [21]. These protective responses are characterised by elevated production of parasite-specific antibodies, including a strong IgE response [22].

Importantly, immunity is *Schistosoma* species-specific, meaning that protection against one species does not necessarily confer protection against others. Different schistosome species can trigger distinct patterns of immune responses in their hosts [23, 24].

Immune responses play a key role not only in protection against infection, but also in disease development. The formation of granulomas around deposited eggs represents a critical immune-driven response that aims to contain the eggs and limit tissue damage. However, while granulomas help isolate the eggs and their secreted products, excessive or prolonged granulomatous inflammation can lead to fibrosis, organ dysfunction, and other *Schistosoma*-related morbidities [1].

1.3 Behavioural Aspects of Transmission

While the complete schistosome life cycle for *S. japonicum* has been described as early as 1913 [26], schistosome biology alone remains insufficient to characterise local pathogen transmission dynamics. A study in Senegal, for instance, mapped snail dynamics using remote sensing and malacology in unprecedented detail and included ten granular snail-related variables in statistical models. Nonetheless, the ability of these models in predicting *S. haematobium* reinfection of individuals was limited as the marginal R^2 was 0.36 [27]. Behavioural factors which vary between villages, households, genders and based on age, could potentially play a

key role in explaining differences in prevalence across geographies and the existence of prevalence hotspot villages [28, 29]. Within the natural schistosome life cycle, behavioural factors intervene at two stages: during human water contact and during contamination (Fig. 1.1).

1.3.1 Human Water Contact Behaviour

In this thesis, schistosome exposure is defined as having water contact at sites infested with cercariae. The type of water contact behaviours that people in endemic areas engage in, such as swimming, fishing, bathing, washing, or laundry, play a critical role in determining infection likelihood as different activities typically have different frequencies and durations and different degrees of immersion [30–33]. Activities may also be done at different times of the day and such variation also implies differences in infection likelihood as there are diurnal patterns in cercarial shedding [34]. It is well-known that contact behaviours are highly variable across small spatial scales, with differences observed in water contact duration, frequency, and

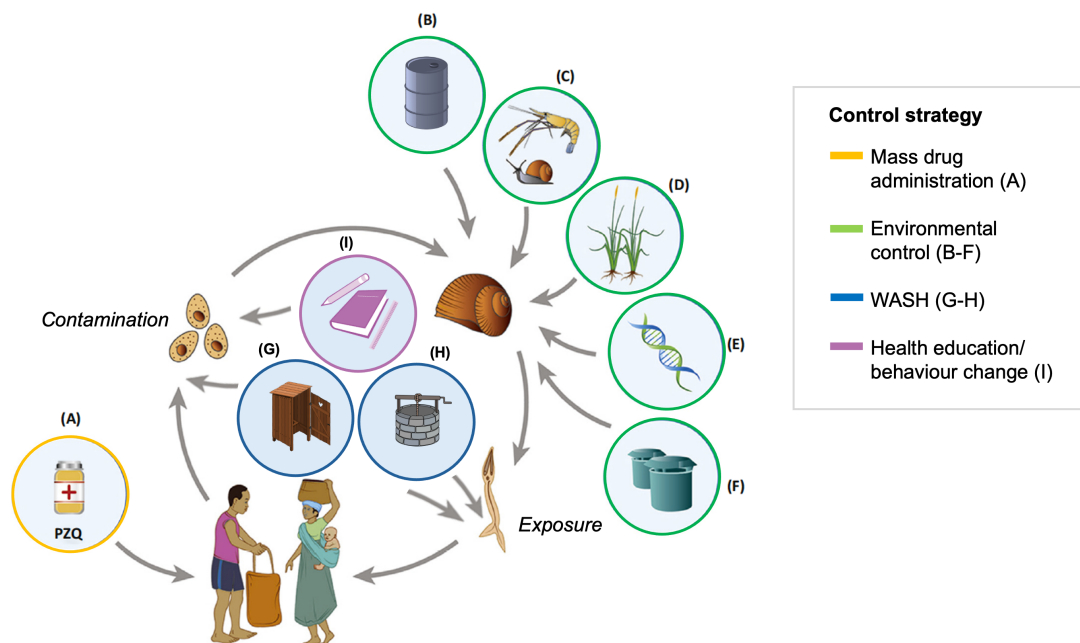


Figure 1.1. The *Schistosoma* sp. lifecycle and control strategies. Control strategy (A) represents mass drug administration (MDA) using praziquantel. (B–F) are environmental control strategies to reduce snail presence: (B) chemical molluscicides, (C) natural predators, (D) habitat modification, (E) creative technologies such as gene drive, (F) traps or repellents and other “out of the box” strategies. (G–H) are water, sanitation, and hygiene (WASH) interventions: (G) provision of safe sanitation infrastructure, (H) provision of safe water supplies. (I) represents health education/behaviour change. This figure was adapted from [25] and reproduced with publisher permission.

activities across households, villages, age groups, genders, and occupational groups [30, 35–40]. Mathematical models have demonstrated that communities with more heterogeneous levels of water contact exhibit higher transmission levels compared to communities with homogeneous water contact patterns, all else being equal [41]. Knowledge of water contact behaviour is crucial for understanding community transmission and individual-level infection risks.

1.3.2 Human Behaviour and Water Site Contamination

Environmental contamination by humans via open defecation, urination, or the lack of safe wastewater disposal enables schistosome eggs to enter the environment and infect competent freshwater snails. In rural communities lacking access to safe WASH, defecation or urination may occur directly at water sites, including sites designated as “public toilets”, or in areas where excreta may be washed into waterbodies [42, 43]. Contamination is a key driver of both local and long-range transmission. For instance, the establishment of a new transmission locus of *S. haematobium* at the Cavu River in Corsica, France, in 2015 was likely due to open urination by infected individuals from Senegal [44]. The primary strategy to reduce contamination of freshwater with human excreta is building safe sanitation infrastructure. As noted by Grimes et al. [45], despite the importance of building safe sanitation infrastructure, preventing contamination may be a less effective control strategy than reducing human water contact. In a study in Ethiopia across 1645 schools, water contact was significantly associated with *S. mansoni* infection while sanitation was not [46]. A key difficulty for sanitation is that near-universal coverage may be required due to the non-linear asexual reproduction of cercariae; one *S. mansoni* egg can produce 340 cercariae and an average infected snail sheds 16,000 cercariae [47]. Thus, local transmission will continue even if sanitation reduces the number of schistosome eggs released into the environment [45].

1.3.3 Human Mobility

Human mobility may influence both local and long-range transmission [48, 49]. Mobile individuals may have increased water contact and visit multiple water sites, increasing their infection risk [50, 51]. However, the opposite has also been reported especially as temporary migration

to urban areas may reduce water contact [49, 52]. Heavily infected, mobile individuals, such as migrant fishing populations, could play a significant role in long-range transmission, potentially introducing the parasite to new geographic areas [53, 54]. These populations may also (re-)introduce *Schistosoma* parasites into their home villages if infections are acquired elsewhere and local water sites are subsequently contaminated.

1.4 Relevance of Human Behavioural Factors for Schistosomiasis Control Interventions and Policy

With no vaccine currently available [55], mass drug administration (MDA) using praziquantel has been adopted as the primary control strategy by the World Health Organization (WHO) [56]. However, praziquantel does not prevent reinfection, and past MDA campaigns have encountered challenges, including low treatment coverage and marginalised households being missed [57]. Concerns have also been raised regarding the potential development of drug resistance with repeated MDA [57, 58]. To achieve the 2030 WHO roadmap targets for neglected tropical diseases [59], MDA must be complemented by additional control interventions, such as WASH provision, environmental control or modification, or behaviour change [60–63].

The efficacy of non-pharmaceutical components within multifaceted control programmes depends on their ability to modify environmental risk or human behaviour (Fig. 1.1). Changes in human behaviour may be achieved through health education, modifications to the social or physical environment, or the implementation of incentives [63]. However, what the most effective approaches to behaviour change are remains open to debate. Among behaviour change interventions in a recent systematic review by Torres-Vitolas et al., 71% relied on self-reported measures to capture changes in water contact behaviour, and only 48% of studies randomly assigned participants to interventions [63]. The most rigorous evaluation of the impacts of behaviour change on schistosome infection was a randomised controlled trial (RCT) conducted in Zanzibar, a near-elimination setting [61]. Compared to the MDA-only arm, five-year *S. haematobium* infection prevalence was not significantly lower in the multifaceted control intervention arms (odds ratio (OR) behaviour change and MDA = 1.31; 95% confidence interval (CI): 0.6–2.9. OR snail control and MDA = 1.21; 95% CI: 0.6–2.7). Knopp et al. chose to focus

their behaviour change intervention on school-aged children since they believed that children swimming, playing, washing laundry, or urinating in local waterbodies were major sources of local transmission [64]. This focus on behaviour change among children is emblematic of the broader literature.

Among the 26 studies in Torres-Vitolas et al. that aimed to modify exposure behaviours, over half (54%) focused on children. It has been argued that laundry, bathing, and recreational swimming are among the activities with the highest exposure, while other activities, such as collecting drinking water, have been deemed unimportant [45]. Yet, systematic evidence on which activities are associated with the highest infection risk and which demographic groups have the highest exposure levels is currently missing.

The ability of WASH interventions to reduce schistosome transmission depends on their ability to modify human behaviour. WASH accessibility cannot necessarily be equated with WASH use but usage is often not adequately measured in studies with infection as the primary outcome. For instance, a meta-analysis by Grimes et al. found that access to safe water supplies and safe sanitation was associated with significantly lower odds of infection (OR safe water supplies = 0.53; 95% (CI): 0.47–0.61. OR safe sanitation (*S. mansoni*) = 0.59, 95% CI: 0.47–0.73). OR safe sanitation *S. haematobium* = 0.69, 95% CI: 0.57–0.84). The authors were unable to distinguish between access and use of safe WASH as such data was not reported in the included studies [45]. In Knopp et al., the RCT arm with behaviour change and MDA also included the construction of washing basins and latrines to reduce water contact and contamination. The authors note that while washing basins were widely used, latrines were not [61], suggesting that WASH may not have averted all risky behaviours, which could be an explanation for why the intervention arm did not outperform the MDA-only arm.

1.5 Relevance of Behavioural Factors for Mathematical Transmission Models

Mathematical transmission models are important tools to guide schistosomiasis control efforts, including those by the WHO [56]. Such models require information on human behaviour as inputs for parameterisation. In terms of water contact, this is typically incorporated as an age-

stratified parameter. For instance, a stochastic individual-based transmission model SCHISTOX uses an age-adjusted water contact frequency and a cercariae uptake rate to model exposure [65]. These models operate under the assumption that age-specific infection levels are a non-linear function of cumulative water contact up to that age and the survival of adult schistosomes within the host [65, 66].

Human mobility is typically not incorporated into transmission models. In SCHISTOX, all individuals are part of the same population with a single shared environmental reservoir of infection. There is a need for more realistic models that account for heterogeneous human-environmental interactions [67]. This could be achieved, for instance, by building metapopulation models where people and snails interact in discrete compartments. This would allow for different rates of water contact across compartments and, potentially, incorporating mobility or connectivity between compartments [68]. While the relevance of mobility for transmission is acknowledged, few examples of mathematical models explicitly incorporating measures of mobility exist [49, 69]. Thus far, modelling has been limited by empirical data on mobility and has relied on proxy measures, such as mobility from call detail records, which are coarse in rural areas [48, 49]. Some work on mobility is done entirely without the use of mobility measures. For instance, Hardwick et al. relied on human settlement data to model hypothetical movement between settlements and a focal transmission point [69]. Therefore, granular empirical data on human water contact and mobility could provide insights into how such behaviours could be best integrated into future transmission models.

1.6 Research Questions

This thesis asked the following overarching research question: What are the roles of human water contact, mobility, and safe water supplies in the epidemiology of *S. mansoni* in rural Uganda?

The following questions were answered in each chapter:

- **Chapter 2:** What is the average effect of water contact frequency, duration, and activity type on the likelihood of schistosome infection?

- **Chapter 3:** What are the major human-environmental determinants of water contact, and at which level (individual, household, or village) is the influence of those determinants? What are (if any) the shared dimensions of water contact determinants with infection determinants?
- **Chapter 4:** Are spatial decay models able to predict fine-scale water site usage and crowding out via tap/borehole usage? What is the role of human mobility in water site usage behaviour?

1.7 Thesis Aims and Content

The aim of this thesis is to advance our understanding of the epidemiology of *S. mansoni* in rural communities across all individuals aged five years and older (not only school-aged children). The three results chapters presented here employ data-driven methods to analyse the influence of current water contact, safe water supplies, and human mobility in a large-scale study (SchistoTrack) in three distinct rural districts in Uganda [70]. The contributions of the main results chapters are as follows:

- **Chapter 2** provides a global overview of the literature on water contact and schistosome infection in endemic settings and presents results from a systematic review and meta-analysis of 101 studies on the associations between having any water contact, water contact duration, and frequency with schistosome infection.
- **Chapter 3** draws upon baseline data on 2867 individuals from SchistoTrack. This analysis investigates age-dependent trends in water contact and infection, water contact-infection associations, and uses Bayesian variable selection to separately identify the key human-environmental drivers of water contact and infection status.
- **Chapter 4** uses wearable GPS logger data on 452 individuals (as a subset of the SchistoTrack participants) to predict individual-level water site and tap/borehole usage with spatial decay models. This chapter further investigates the existence of crowding-out of water contact by tap/borehole infrastructure and provides population-wide predictions of water site usage patterns.

1.8 Publications Arising from this Thesis

This thesis has resulted in two publications. Chapter 2 has been published as:

1. **Reitzug F**, Ledien J, and Chami GF. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. *PLOS Neglected Tropical Diseases* 2023 Jun; 17:e0011377. doi: 10.1371/journal.pntd.0011377.

Chapter 3 has been published as:

1. **Reitzug F**, Kabatereine NB, Byaruhanga AM, Besigye F, Nabatte B, and Chami GF. Current water contact and *Schistosoma mansoni* infection have distinct determinants: A data-driven population-based study in rural Uganda. *Nature Communications* 2024 Nov; 15. doi: 10.1038/s41467-024-53519-4.

1.9 Contributions and Funding

All statistical analyses, original drafts, as well as publications arising from this thesis are my own work. I want to explicitly acknowledge the important contributions that others have made. Goylette F. Chami (GFC) supervised all chapters. Renaud Lambiotte (RL) supervised primarily Chapter 4. Funding for fieldwork has been provided by GFC as well as by the Nuffield Department of Population Health (NDPH). Julia Ledien (JL) double-screened 10% of studies for my systematic review in Chapter 2 and provided comments on an early draft version. For Chapter 3, I received feedback from Narcis B. Kabatereine, (NBK), Anatol M. Byaruhanga (AMB), Fred Besigye (FB), and Betty Nabatte (BN). Prathyush Sambaturu (PS) gave helpful suggestions during the initial stages of Chapter 4. Many individuals supported the fieldwork that enabled the data analyses presented in Chapters 3–4. The Ugandan field team collected the household and clinical data with supervision from GFC, NKB, BN, and Benjamin Tintika (BT). Data management and field coordination support was provided by GFC and Max Eyre (ME). GFC, Olympia Lamberti (OL), ME, and GFC programmed, tested and revised the baseline questionnaires. ME and Farina Shabaan (FS) helped to set up the GPS loggers and with testing in Oxford. GFC and ME supported collecting the village infrastructure data. Post-fieldwork, Adam Fowler (AF) helped with initial attempts to retrieve data from damaged GPS loggers. Keiran Foster (KF) and Joseph Brown (JB) from the University of Oxford Materials Science

Department managed to pull GPS logger data from damaged devices by re-soldering chips to a working logger. All analyses for Chapters 3–4 were conducted on the Biomedical Research Computing (BMRC) facility cloud where Geoffrey Ferrari (GF) provided instrumental help with running computing jobs and installing software.

On a stylistic note; given the very substantial teamwork involved in collecting the SchistoTrack study data used in Chapters 3–4, I will use “we” in those chapters to adequately reflect that this work would not have been possible without the involvement of those acknowledged here.

1.10 Ethics Approval

SchistoTrack 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 the Uganda National Council of Science and Technology (UNCST HS 1664ES).

References

- McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, and Zhou XN. Schistosomiasis. *Nature Reviews Disease Primers* 2018 Aug; 4. doi: 10.1038/s41572-018-0013-8
- Naghavi M et al. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet* 2024 May; 403:2100–32. doi: 10.1016/s0140-6736(24)00367-2
- Li Q, Li YL, Guo SY, Li SZ, Wang Q, Lin WN, Zhang LJ, Li SZ, Zhou XN, and Xu J. Global trends of schistosomiasis burden from 1990 to 2021 across 204 countries and territories: Findings from GBD 2021 study. *Acta Tropica* 2025 Jan; 261:107504. doi: 10.1016/j.actatropica.2024.107504
- Steinmann P, Keiser J, Bos R, Tanner M, and Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases* 2006 Jul; 6:411–25. doi: 10.1016/S1473-3099(06)70521-7
- Lai YS, Biedermann P, Ekpo UF, Garba A, Mathieu E, Midzi N, Mwinzi P, N’Goran EK, Raso G, Assaré RK, Sacko M, Schur N, Talla I, Tchuenté LAT, Touré S, Winkler MS, Utzinger J, and Vounatsou P. Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. *The Lancet Infectious Diseases* 2015 Aug; 15:927–40. doi: 10.1016/s1473-3099(15)00066-3
- King CH. Parasites and poverty: The case of schistosomiasis. *Acta Tropica* 2010 Feb; 113:95–104. doi: 10.1016/j.actatropica.2009.11.012
- Assis AMO, Barreto ML, Prado MS, Reis MG, Parraga IM, and Blanton RE. *Schistosoma mansoni* infection and nutritional status in schoolchildren: a randomized, double-blind trial in northeastern Brazil. *The American Journal of Clinical Nutrition* 1998 Dec; 68:1247–53. doi: 10.1093/ajcn/68.6.1247
- Chami GF, Fenwick A, Bulte E, Kontoleon AA, Kabatereine NB, Tukahebwa EM, and Dunne DW. Influence of *Schistosoma mansoni* and Hookworm Infection Intensities on Anaemia in Ugandan Villages. *PLoS Neglected Tropical Diseases* 2015 Oct; 9:e0004193. doi: 10.1371/journal.pntd.0004193

CHAPTER 1. INTRODUCTION

9. Ezeamama AE, Bustinduy AL, Nkwata AK, Martinez L, Pabalan N, Boivin MJ, and King CH. Cognitive deficits and educational loss in children with schistosome infection—A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2018 Jan; 12:e0005524. doi: 10.1371/journal.pntd.0005524
10. Vennervald BJ and Polman K. Helminths and malignancy. *Parasite Immunology* 2009 Oct; 31:686–96. doi: 10.1111/j.1365-3024.2009.01163.x
11. Lenk EJ, Redekop WK, Luyendijk M, Rijnsburger AJ, and Severens JL. Productivity Loss Related to Neglected Tropical Diseases Eligible for Preventive Chemotherapy: A Systematic Literature Review. *PLoS Neglected Tropical Diseases* 2016 Feb; 10:e0004397. doi: 10.1371/journal.pntd.0004397
12. Rinaldo D, Perez-Saez J, Vounatsou P, Utzinger J, and Arcand JL. The economic impact of schistosomiasis. *Infectious Diseases of Poverty* 2021 Dec; 10. doi: 10.1186/s40249-021-00919-z
13. Bonds MH, Keenan DC, Rohani P, and Sachs JD. Poverty trap formed by the ecology of infectious diseases. *Proceedings of the Royal Society B: Biological Sciences* 2009 Dec; 277:1185–92. doi: 10.1098/rspb.2009.1778
14. Miguel E and Kremer M. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 2004 Jan; 72:159–217. doi: 10.1111/j.1468-0262.2004.00481.x
15. Stensgaard AS, Utzinger J, Vounatsou P, Hürlimann E, Schur N, Saarnak CF, Simoonga C, Mubita P, Kabatereine NB, Tchuem Tchuenté LA, Rahbek C, and Kristensen TK. Large-scale determinants of intestinal schistosomiasis and intermediate host snail distribution across Africa: does climate matter? *Acta Tropica* 2013 Nov; 128:378–90. doi: 10.1016/j.actatropica.2011.11.010
16. Hailegebriel T, Nibret E, and Munshea A. Prevalence of *Schistosoma mansoni* and *S. haematobium* in Snail Intermediate Hosts in Africa: A Systematic Review and Meta-analysis. *Journal of Tropical Medicine* 2020 Sep; 2020:1–18. doi: 10.1155/2020/8850840
17. Clegg J. In vitro cultivation of *Schistosoma mansoni*. *Experimental Parasitology* 1965 Apr; 16:133–47. doi: 10.1016/0014-4894(65)90037-8
18. Lieshout L van, Polderman AM, Visser LG, Verwey JJ, and Deelder AM. Detection of the circulating antigens CAA and CCA in a group of Dutch travellers with acute schistosomiasis. *Tropical Medicine & International Health* 1997 Jun; 2:551–7. doi: 10.1046/j.1365-3156.1997.d01-324.x
19. Arnold BF, Kanyi H, Njenga SM, Rawago FO, Priest JW, Secor WE, Lammie PJ, Won KY, and Odiere MR. Fine-scale heterogeneity in *Schistosoma mansoni* force of infection measured through antibody response. *Proceedings of the National Academy of Sciences* 2020 Aug; 117:23174–81. doi: 10.1073/pnas.2008951117
20. Woolhouse MEJ, Taylor P, Matanhire D, and Chandiwana SK. Acquired immunity and epidemiology of *Schistosoma haematobium*. *Nature* 1991 Jun; 351:757–9. doi: 10.1038/351757a0
21. Black CL, Mwinzi PNM, Muok EMO, Abudho B, Fitzsimmons CM, Dunne DW, Karanja DMS, Secor WE, and Colley DG. Influence of Exposure History on the Immunology and Development of Resistance to Human Schistosomiasis *Mansoni*. *PLoS Neglected Tropical Diseases* 2010 Mar; 4. Ed. by Bethony JM:e637. doi: 10.1371/journal.pntd.0000637
22. Hagan P, Blumenthal UJ, Dunn D, Simpson AJG, and Wilkins HA. Human IgE, IgG4 and resistance to reinfection with *Schistosoma haematobium*. *Nature* 1991 Jan; 349:243–5. doi: 10.1038/349243a0
23. Oettle RC, Dickinson HA, Fitzsimmons CM, Sacko M, Tukahebwa EM, Chalmers IW, and Wilson S. Protective human IgE responses are promoted by comparable life-cycle dependent Tegument Allergen-Like expression in *Schistosoma haematobium* and *Schistosoma mansoni* infection. *PLOS Pathogens* 2023 May; 19. Ed. by Loke P:e1011037. doi: 10.1371/journal.ppat.1011037
24. Oettle RC. Human antibody-mediated immunity against infection and parasite fecundity in *Schistosoma mansoni* and *Schistosoma haematobium* infection. PhD thesis. University of Cambridge, 2020 Jul
25. Sokolow SH, Wood CL, Jones IJ, Lafferty KD, Kuris AM, Hsieh MH, and De Leo GA. To Reduce the Global Burden of Human Schistosomiasis, Use ‘Old Fashioned’ Snail Control. *Trends in Parasitology* 2018 Jan; 34:23–40. doi: 10.1016/j.pt.2017.10.002
26. Jamieson BL. *Schistosoma* Biology, Pathology and Control. CRC Press, 2017 Jan. doi: 10.1201/9781315368900

27. Wood CL, Sokolow SH, Jones IJ, Chamberlin AJ, Lafferty KD, Kuris AM, Jocque M, Hopkins S, Adams G, Buck JC, Lund AJ, Garcia-Vedrenne AE, Fiorenza E, Rohr JR, Allan F, Webster B, Rabone M, Webster JP, Bandagny L, Ndione R, Senghor S, Schacht AM, Jouanard N, Riveau G, and De Leo GA. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. *Proceedings of the National Academy of Sciences* 2019 Oct; 116:23182–91. doi: 10.1073/pnas.1903698116
28. Woolhouse M, Watts C, and Chandiwana S. Heterogeneities in transmission rates and the epidemiology of schistosome infection. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 1991; 245:109–14. doi: 10.1098/rspb.1991.0095
29. Singer BJ, Coulibaly JT, Park HJ, Andrews JR, Bogoch II, and Lo NC. Development of prediction models to identify hotspots of schistosomiasis in endemic regions to guide mass drug administration. *Proceedings of the National Academy of Sciences* 2024 Jan; 121. doi: 10.1073/pnas.2315463120
30. Sow S, Vlas SJ de, Stelma F, Vereecken K, Gryseels B, and Polman K. The contribution of water contact behavior to the high *Schistosoma mansoni* Infection rates observed in the Senegal River Basin. *BMC Infectious Diseases* 2011 Dec; 11:198. doi: 10.1186/1471-2334-11-198
31. Dalton PR and Pole D. Water-contact patterns in relation to *Schistosoma haematobium* infection. English. *Bulletin of the World Health Organization* 1978; 56:417–26
32. Lund AJ, Sokolow SH, Jones IJ, Wood CL, Ali S, Chamberlin A, Sy AB, Sam MM, Jouanard N, Schacht AM, Senghor S, Fall A, Ndione R, Riveau G, De Leo GA, and Lopez-Carr D. Exposure, hazard, and vulnerability and their contribution to *Schistosoma haematobium* re-infection in northern Senegal. *The Lancet Planetary Health* 2021 Apr; 5:S10. doi: 10.1016/S2542-5196(21)00094-2
33. Chandiwana SK. Community water-contact patterns and the transmission of *Schistosoma haematobium* in the highveld region of Zimbabwe. *Social Science & Medicine* 1987 Jan; 25:495–505. doi: 10.1016/0277-9536(87)90173-0
34. Sidibé B, Agniwo P, Diakité A, Savassi BAEoS, Doumbo SN, Akplogan A, Guindo H, Ibikounlé M, Dembélé L, Djimde A, Boissier J, and Dabo A. Human-water interactions associated to cercarial emergence pattern and their influences on urinary schistosomiasis transmission in two endemic areas in Mali. *Infectious Diseases of Poverty* 2024 Aug; 13. doi: 10.1186/s40249-024-01229-w
35. Meurs L, Mbow M, Boon N, Broeck F van den, Vereecken K, Dièye TN, Abatih E, Huyse T, Mboup S, and Polman K. Micro-Geographical Heterogeneity in *Schistosoma mansoni* and *S. haematobium* Infection and Morbidity in a Co-Endemic Community in Northern Senegal. *PLoS Neglected Tropical Diseases* 2013 Dec; 7:e2608. doi: 10.1371/journal.pntd.0002608
36. Farooq M and Mallah MB. The behavioural pattern of social and religious water-contact activities in the Egypt-49 bilharziasis project area. *Bulletin of the World Health Organization* 1966; 35:377–87
37. Fulford AJC, Ouma JH, Kariuki HC, Thiongo FW, Klumpp R, Kloos H, Sturrock RF, and Butterworth AE. Water contact observations in Kenyan communities endemic for schistosomiasis: methodology and patterns of behaviour. *Parasitology* 1996 Sep; 113:223–41. doi: 10.1017/S0031182000082007
38. Butterworth AE, Dalton PR, and Dunne DW. Immunity after treatment of human schistosomiasis mansoni. I. Study design, pretreatment observations and the results of treatment. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1984; 78:108–23. doi: 10.1016/0035-9203(84)90190-1
39. De Moira AP, Fulford AJC, Kabatereine NB, Kazibwe F, Ouma JH, Dunne DW, and Booth M. Microgeographical and tribal variations in water contact and *Schistosoma mansoni* exposure within a Ugandan fishing community: Water contact and *Schistosoma mansoni* in Uganda. *Tropical Medicine & International Health* 2007 Jun; 12:724–35. doi: 10.1111/j.1365-3156.2007.01842.x
40. Pinot de Moira A, Fulford AJC, Kabatereine NB, Ouma JH, Booth M, and Dunne DW. Analysis of Complex Patterns of Human Exposure and Immunity to *Schistosomiasis mansoni*: The Influence of Age, Sex, Ethnicity and IgE. *PLoS Neglected Tropical Diseases* 2010 Sep; 4:e820. doi: 10.1371/journal.pntd.0000820
41. Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, Hagan P, Hii JLK, Ndhlovu PD, Quinnell RJ, Watts CH, Chandiwana SK, and Anderson RM. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proceedings of the National Academy of Sciences* 1997 Jan; 94:338–42. doi: 10.1073/pnas.94.1.338
42. Sow S, Polman K, Vereecken K, Vercruysse J, Gryseels B, and Vlas SJ de. The role of hygienic bathing after defecation in the transmission of *Schistosoma mansoni*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2008 Jun; 102:542–7. doi: 10.1016/j.trstmh.2008.02.016

43. Lund AJ, Sam MM, Sy AB, Sow OW, Ali S, Sokolow SH, Bereknayi Merrell S, Bruce J, Jouanard N, Senghor S, Riveau G, Lopez-Carr D, and De Leo GA. Unavoidable Risks: Local Perspectives on Water Contact Behavior and Implications for Schistosomiasis Control in an Agricultural Region of Northern Senegal. *The American Journal of Tropical Medicine and Hygiene* 2019 Oct; 101:837–47. doi: 10.4269/ajtmh.19-0099
44. Berry A, Fillaux J, Martin-Blondel G, Boissier J, Iriart X, Marchou B, Magnaval JF, and Delobel P. Evidence for a permanent presence of schistosomiasis in Corsica, France, 2015. *Eurosurveillance* 2016 Jan; 21. doi: 10.2807/1560-7917.es.2016.21.1.30100
45. Grimes JE, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasites & Vectors* 2015 Dec; 8:156. doi: 10.1186/s13071-015-0766-9
46. Grimes JET, Tadesse G, Mekete K, Wuletaw Y, Gebretsadik A, French MD, Harrison WE, Drake LJ, Gardiner IA, Yard E, and Templeton MR. School Water, Sanitation, and Hygiene, Soil-Transmitted Helminths, and Schistosomes: National Mapping in Ethiopia. *PLoS Neglected Tropical Diseases* 2016 Mar; 10:e0004515. doi: 10.1371/journal.pntd.0004515
47. Kloetzel K and Corr  a LR dos. Some quantitative data on the laboratory and field infection of *Biomphalaria glabrata* from Jabuticatubas (State of Minas Gerais, Brasil) by miracida of *Schistosoma mansoni*. *Revista do Instituto de Medicina Tropical de S  o Paulo* 1968 Apr; 10:59–69
48. Ciddio M, Mari L, Sokolow SH, De Leo GA, Casagrandi R, and Gatto M. The spatial spread of schistosomiasis: A multidimensional network model applied to Saint-Louis region, Senegal. *Advances in Water Resources* 2017 Oct; 108:406–15. doi: 10.1016/j.advwatres.2016.10.012
49. Mari L, Gatto M, Ciddio M, Dia ED, Sokolow SH, De Leo GA, and Casagrandi R. Big-data-driven modeling unveils country-wide drivers of endemic schistosomiasis. *Scientific Reports* 2017 Mar; 7. doi: 10.1038/s41598-017-00493-1
50. Cetron MS, Chitsulo L, Sullivan JJ, Pilcher J, Wilson M, Noh J, Tsang VC, Hightower AW, and Addiss DG. Schistosomiasis in Lake Malawi. *The Lancet* 1996 Nov; 348:1274–8. doi: 10.1016/s0140-6736(96)01511-5
51. Bella H, C. Marshall T de, Omer A, and Vaughan J. Migrant workers and schistosomiasis in the Gezira, Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1980 Jan; 74:36–9. doi: 10.1016/0035-9203(80)90008-5
52. Buchwald AG, Grover E, Van Dyke J, Kechris K, Lu D, Liu Y, Zhong B, and Carlton EJ. Human Mobility Associated With Risk of *Schistosoma japonicum* Infection in Sichuan, China. *American Journal of Epidemiology* 2021 Jan. doi: 10.1093/aje/kwaa292
53. Perez-Saez J, Mari L, Bertuzzo E, Casagrandi R, Sokolow SH, De Leo GA, Mande T, Ceperley N, Froehlich JM, Sou M, Karambiri H, Yacouba H, Maiga A, Gatto M, and Rinaldo A. 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 Neglected Tropical Diseases* 2015 Oct; 9. Ed. by Liang S:e0004127. doi: 10.1371/journal.pntd.0004127
54. Walz Y, Wegmann M, Dech S, Raso G, and Utzinger J. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. *Parasites & Vectors* 2015 Dec; 8:163. doi: 10.1186/s13071-015-0732-6
55. Yamey G, McDade KK, Anderson RM, Bartsch SM, Bottazzi ME, Diemert D, Hotez PJ, Lee BY, McManus D, Molehin AJ, Roestenberg M, Rollinson D, Siddiqui AA, Tendler M, Webster JP, You H, Zellweger RM, and Marshall C. Vaccine value profile for schistosomiasis. *Vaccine* 2024 Jul :126020. doi: 10.1016/j.vaccine.2024.05.068
56. World Health Organization. WHO guideline on control and elimination of human schistosomiasis. Geneva: World Health Organization, 2022
57. Chami GF, Kontoleon AA, Bulte E, Fenwick A, Kabatereine NB, Tukahebwa EM, and Dunne DW. Profiling Nonrecipients of Mass Drug Administration for Schistosomiasis and Hookworm Infections: A Comprehensive Analysis of Praziquantel and Albendazole Coverage in Community-Directed Treatment in Uganda. *Clinical Infectious Diseases* 2016 Sep; 62:200–7. doi: 10.1093/cid/civ829

58. Crellen T, Walker M, Lamberton PHL, Kabatereine NB, Tukahebwa EM, Cotton JA, and Webster JP. Reduced Efficacy of Praziquantel Against *Schistosoma mansoni* Is Associated With Multiple Rounds of Mass Drug Administration. *Clinical Infectious Diseases* 2016 Jul; 63:1151–9. doi: 10.1093/cid/ciw506
59. Organization WH. Ending the neglect to attain the Sustainable Development Goals: a road map for neglected tropical diseases 2021–2030. World Health Organization, 2020
60. Hoover CM, Sokolow SH, Kemp J, Sanchirico JN, Lund AJ, Jones IJ, Higginson T, Riveau G, Savaya A, Coyle S, Wood CL, Micheli F, Casagrandi R, Mari L, Gatto M, Rinaldo A, Perez-Saez J, Rohr JR, Sagi A, Remais JV, and De Leo GA. Modelled effects of prawn aquaculture on poverty alleviation and schistosomiasis control. *Nature Sustainability* 2019 Jul; 2:611–20. doi: 10.1038/s41893-019-0301-7
61. Knopp S, Person B, Ame SM, Ali SM, Hattendorf J, Juma S, Muhsin J, Khamis IS, Mohammed KA, Utzinger J, Hollenberg E, Kabole F, and Rollinson D. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. *The Lancet Global Health* 2019 Aug; 7:e1118–e1129. doi: 10.1016/s2214-109x(19)30189-5
62. Rohr JR, Sack A, Bakhoum S, Barrett CB, Lopez-Carr D, Chamberlin AJ, Civitello DJ, Diatta C, Doruska MJ, De Leo GA, Haggerty CJE, Jones IJ, Jouanard N, Lund AJ, Ly AT, Ndione RA, Remais JV, Riveau G, Schacht AM, Seck M, Senghor S, Sokolow SH, and Wolfe C. A planetary health innovation for disease, food and water challenges in Africa. *Nature* 2023 Jul; 619:782–7. doi: 10.1038/s41586-023-06313-z
63. Torres-Vitolas CA, Trienekens SCM, Zaadnoordijk W, and Gouvras AN. Behaviour change interventions for the control and elimination of schistosomiasis: A systematic review of evidence from low- and middle-income countries. *PLoS Neglected Tropical Diseases* 2023 May; 17:e0011315. doi: 10.1371/journal.pntd.0011315
64. Knopp S, Mohammed KA, Ali SM, Khamis IS, Ame SM, Albonico M, Gouvras A, Fenwick A, Savioli L, Colley DG, Utzinger J, Person B, and Rollinson D. Study and implementation of urogenital schistosomiasis elimination in Zanzibar (Unguja and Pemba islands) using an integrated multidisciplinary approach. *BMC Public Health* 2012 Oct; 12. doi: 10.1186/1471-2458-12-930
65. Graham M, Ayabina D, Lucas TC, Collyer BS, Medley GF, Hollingsworth TD, and Toor J. SCHISTOX: An individual based model for the epidemiology and control of schistosomiasis. *Infectious Disease Modelling* 2021; 6:438–47. doi: 10.1016/j.idm.2021.01.010
66. Anderson R, Turner H, Farrell S, and Truscott J. Studies of the Transmission Dynamics, Mathematical Model Development and the Control of Schistosome Parasites by Mass Drug Administration in Human Communities. *Mathematical Models for Neglected Tropical Diseases - Essential Tools for Control and Elimination, Part B*. Elsevier, 2016 :199–246. doi: 10.1016/bs.apar.2016.06.003
67. Gurarie D, Lo NC, Ndeffo-Mbah ML, Durham DP, and King CH. The human-snail transmission environment shapes long term schistosomiasis control outcomes: Implications for improving the accuracy of predictive modeling. *PLoS Neglected Tropical Diseases* 2018 May; 12:e0006514. doi: 10.1371/journal.pntd.0006514
68. Fenton A, Fairbairn JP, Norman R, and Hudson PJ. Parasite transmission: reconciling theory and reality. *Journal of Animal Ecology* 2002 Sep; 71:893–905. doi: 10.1046/j.1365-2656.2002.00656.x
69. Hardwick RJ, Vegvari C, Collyer B, Truscott JE, and Anderson RM. Spatial scales in human movement between reservoirs of infection. *Journal of Theoretical Biology* 2021 Sep; 524:110726. doi: 10.1016/j.jtbi.2021.110726
70. Nuffield Department of Population Health, University of Oxford. SchistoTrack: a prospective multimorbidity cohort. <https://www.bdi.ox.ac.uk/research/schistotrack>. [Accessed on: 2024 Apr 16]

Chapter 2

Associations of Water Contact with Schistosome Infection Risk: A Systematic Review and Meta-Analysis

This chapter has been published as:

1. **Reitzug F**, Ledien J, and Chami GF. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2023 Jun; 17:e0011377. doi: 10.1371/journal.pntd.0011377.

Author contributions: Conceptualisation: FR and GFC. Data curation: FR and JL. Formal analysis: FR. Funding acquisition: GFC. Investigation: FR. Methodology: FR and GFC. Project administration: GFC. Resources: GFC. Software: GFC. Supervision: GFC. Validation: FR and JL. Visualisation: FR. Writing – original draft: FR. Writing – review & editing: FR, JL, and GFC.

2.1 Background

In this chapter, a systematic review and meta-analysis was conducted to quantify associations between having any water contact and water contact duration and frequency with schistosome infection likelihood.

The motivation for this review was the substantial heterogeneity in such associations reported across the literature. The magnitude of associations between water contact and schistosome infection has varied across studies, and some studies did not find any significant association. For instance, Costa et al. [1] conducted a survey in rural south-eastern Brazil and found that water contact was significantly more strongly associated with the odds of schistosome infection

in people aged 14 years and younger (odds ratio (OR) 55.8; 95% confidence interval (CI): 27.2–114.6) compared to people aged 15 years and older (OR 6.8; 95% CI: 3.5–13.2). By contrast, a nationally representative cross-sectional survey in Uganda by Exum et al. found that self-reported contact with open freshwater sources was not significantly associated with schistosome infection, even when the effect size was unadjusted (prevalence ratio 1.08; 95% CI: 0.9–1.26) [2].

Past studies have used surveys, diaries, direct water contact observation, and wearable global positioning system (GPS) loggers to measure water contact. Surveys provide cross-sectional measures of current water contact. Self-reported measures have been found to agree relatively poorly with observation-based measures of water contact in studies which compared the two [3, 4]. Direct water contact observation using trained local observers has provided longitudinal measures of water contact within communities. These data have contributed to understanding typical levels of exposure in endemic settings as well as daily, weekly, monthly, and seasonal water contact patterns. Observation studies have been able to document large numbers of water contacts over extended time periods. For instance, Fulford et al. conducted direct water contact observations in seven Kenyan communities spanning three years and recording 67,000 contacts [5]. Sow et al. observed water contacts in a community in northern Senegal for two years, recording 120,000 water contacts [6]. The accuracy of water contact observation has been validated by assessing inter-observer agreement [5, 6]. Limitations of observation are that it is expensive, data processing is time-consuming, and resulting exposure measures have often been used as ecological, not individual-level, measures [5–9]. Among the two above-mentioned observation studies, none reported on associations between water contact and individual-level infection outcomes.

A method that has been made possible by advances in wearable technology (lower cost, smaller size, longer battery life) [10] are water contact measurements using GPS loggers. GPS loggers provide longitudinal measures but so far, the existing studies in Cameroon and Uganda have been restricted to short observation periods (24 hours and 72 hours, respectively) and small samples ($n = 24$, $n = 97$) [4, 11]. Across all water contact measurement tools, no accepted gold standard for exposure measurement exists.

There is a need to systematically analyse associations between water contact and schistosome infection to (1) disentangle the relative role of exposure versus acquired immunity in explaining individuals' likelihood of infection [12, 13], (2) assess whether there is a dose-response relationship between degree of water contact and schistosome infection intensity [6, 14], and (3) examine the extent of the age-dependency of infection [15, 16]. It remains an open question as to what water contact measures are associated with schistosome infection status and intensity. Therefore, a systematic review and meta-analysis was performed to address the following question: What is the average effect of water contact frequency, duration, and activity type on the likelihood of schistosome infection?

2.2 Methods

2.2.1 Protocol and Search Strategy

The protocol was registered on 13 May, 2022 with PROSPERO (CRD42022333680) [17]. The systematic review has been reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [18]. The completed checklist is provided as an Online Supplementary S1 Text [19]¹.

Six databases in English were searched on 13 May, 2022: Embase (1974-), MEDLINE (including PubMed, 1946-), Global Health (1973-), Global Index Medicus (1901-), Web of Science Core Collection (1900-), Science Citation Index Expanded (1900-), and the Cochrane Central Register of Controlled Trials (CENTRAL, 1996-). The following search string was employed: “Schistosomiasis” AND (“water” (AND “contact” OR “pattern” OR “duration” OR “frequency” OR “behaviour” OR “exposure”)) AND (“risk” OR “infection” OR “intensity” OR “transmission” OR “odds” OR “likelihood”). A combination of free text search and Medical Subject Heading (MeSH) terms was utilised. Full details of the search strategy are provided in Online Supplementary S2 Text [19].

¹All online materials in this chapter are additionally attached as supplementary thesis materials.

2.2.2 Exposures and Outcome

The Population-Exposure-Comparators-Outcomes (PECO) framework was utilised to formulate the research question [20]. The study population comprised individuals of any age or gender living in areas endemic with a schistosome species. Water contact was categorised into four distinct exposures: having any water contact, water contact duration, water contact frequency, and water contact activities.

- **Having any water contact** was defined as a binary measure, i.e., whether the respondent had any contact with an open freshwater body over a specified recall or observation period.
- **Water contact frequency** was defined as a categorical measure indicating the number of contacts per day, week, or month.
- **Water contact duration** was defined as a categorical measure, indicating the duration per water contact event or over the recall/observation period.
- **Water contact activities** were defined as binary measures of engagement in specific activities (e.g., bathing, swimming, fishing).

Water contact measures from any water contact measurement tool were eligible. The comparators included individuals without any water contact or with lower frequency or duration of contact, depending on the unit of exposure. The primary outcome was the likelihood of schistosome infection (infection status or intensity) based on recognised diagnostics such as microscopy (e.g., Kato–Katz method [21], urine filtration [22]) or antigen/antibody-based methods (e.g., point-of-care circulating cathodic antigen tests [23, 24] or polymerase chain reaction-based detection [25]). Effect measures were ORs or hazard ratios (HRs) with 95% CIs. Wherever available, HRs or adjusted ORs were extracted. Otherwise, unadjusted ORs, or information to convert reported associations to ORs, such as risk ratios or 2×2 tables were used.

2.2.3 Inclusion and Exclusion Criteria

Eligible study designs included cohort studies, cross-sectional studies, case-control studies, before-after studies (reporting baseline infection and exposure), and randomised controlled trials

(reporting infection and exposure at baseline or in control groups). Studies using self-reports or (micro-)haematuria to ascertain infection status were excluded. Only exposure measures with identifiable dimensions (e.g., having any water contact, frequency, duration, activity categories) were included. Exposure indices that collapsed multiple different dimensions of water contact (and sometimes also measures of environmental risk) into a single metric and which could not be mapped onto the four water contact dimensions were excluded.

2.2.4 Screening and Data Extraction

Title, abstract, and full-text screening was performed independently by two reviewers (myself and JL). I extracted all data, with 10% of included studies independently extracted by JL for quality control. For duplicate data from overlapping populations, the study with the larger sample size was retained. If the data to reconstruct measures of association were not reported, associations could not be reconstructed from published data including supplements or required information was not provided by the authors after e-mail contact, the study was excluded.

A structured, pre-tested form was used to extract study characteristics: study design, study location, study setting, local waterbodies, sampling strategy, participant characteristics, inclusion/exclusion criteria, exposure measurements, *Schistosoma* sp., definition of the schistosome infection outcome, diagnostic test, and schistosome infection prevalence in study participants. More details on the data extraction are available in Online Supplementary S3 Text [19]. References of all included studies and a list of excluded studies can be found in Online Tables S1–S2, respectively [19].

2.2.5 Subgroup Variables

Studies were grouped by the following characteristics:

- **Age:** Children (ages 0-14) were further subdivided into pre-school (ages 0-4) and school-age (ages 5-14). Adults were defined as age ≥ 15 years.
- **Endemicity:** Studies were classified using 2022 WHO schistosomiasis control guideline prevalence thresholds [26]: low ($< 10\%$), moderate (10-49%), and high ($\geq 50\%$).

- **Waterbody characteristics:** Stagnant (e.g., lakes, ponds) versus flowing (e.g., rivers, streams) and artificial (e.g., reservoirs) versus natural.
- **Climate:** The local Köppen-Geiger classification was assigned based on study GPS coordinates [27].

Duration was dichotomised (≤ 1 hour or > 1 hour), and frequency was categorised (daily, weekly, monthly). Water contact activities were grouped as domestic, recreational, or occupational. Online Supplementary Table S3 details these groupings [19].

2.2.6 Data Analysis

For all possible water contact-infection association pairs, random-effects meta-analyses with inverse variance weighting following the DerSimonian and Laird procedure were conducted whenever at least three studies reported comparable effect measures [28]. Adjusted effect sizes were preferred because they better account for potential confounding factors in the water contact-infection relationship. In the water contact duration and activity analyses where it was possible for studies to contribute more than one effect size, multilevel meta-analysis was used. Pre-specified subgroup analyses were performed where sufficient data (≥ 3 studies) were available. Between-study heterogeneity was assessed using the I^2 -statistic, Egger's test, and funnel plots. Analysis was implemented using the `meta` and `dmetar` packages in R version 4.2.1 [29–31].

2.2.7 Risk of Bias

Study quality was appraised using a modified version of the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies from the National Institutes of Health (NIH, Bethesda, MD, United States of America) [32]. Risk of bias was categorised as low, moderate, or high (see Online Supplementary S4 Text [19]). The preferred method for assessing the effect of study quality on the results in this chapter was to perform meta-analyses separated by risk of bias. As few studies were rated high quality, there were not enough studies to use this method in frequency, duration, and activity meta-analyses. In this case, influence analysis was used as

a secondary method. To this end, one study at a time was removed from each meta-analysis to quantify the effect of any single study on the results.

2.3 Results

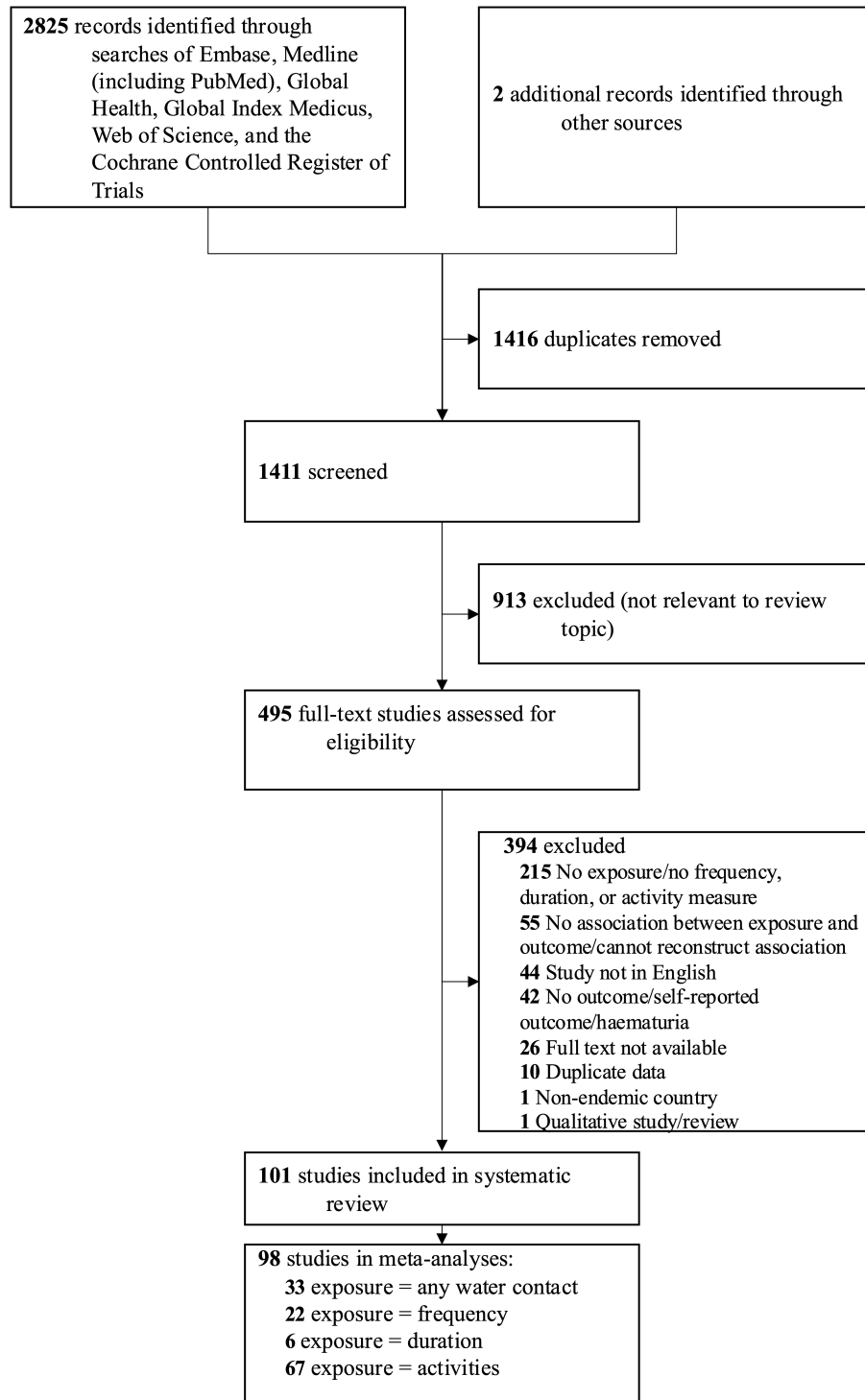


Figure 2.1. Study selection. PRISMA Flowchart depicting the number of records identified, included, and excluded, along with the reasons for exclusion.

2.3.1 Study Characteristics

The search returned 1411 unique titles and abstracts. A total of 495 studies met the inclusion criteria for full-text review, and 101 studies were included (Fig. 2.1). Summary characteristics of all included studies are reported in Table 2.1. Detailed study-level information and full references of all included studies are provided in Online Table S1 [19]. Online Table S2 provides a list of excluded studies and details their exclusion reasons [19]. Of the 101 studies in this systematic review, representing 192,691 participants and reporting on 501 associations between water contact and schistosome infection, 98 studies (97%), with 191,146 participants, provided comparable outcome and exposure measures and were included in at least one full or subgroup meta-analysis.

Table 2.1. Study characteristics summary

Characteristics	N (%)
Study aims	
Measurement of water contact	10 (9.9%)
Other aim	91 (90.1%)
Continent	
Africa	78 (77.2%)
Asia	10 (9.9%)
South America	13 (12.9%)
Locality	
Rural	77 (76.2%)
Urban	4 (4.0%)
Mixed ²	20 (19.8%)
Study setting	
Community	63 (62.4%)
School	31 (30.7%)
Other ³	7 (6.9%)
Population⁴	
PSACs/SACs	33 (32.7%)
PSACs/SACs/Adults	68 (67.3%)
<i>Schistosoma</i> sp.	
<i>S. haematobium</i>	47 (46.5%)
<i>S. mansoni</i>	42 (41.6%)
<i>S. japonicum</i>	5 (5.0%)
Multiple species	7 (6.9%)
Endemicity	

Continued on next page

²Study areas that include both rural and (peri-)urban areas.

³Health clinic or multiple settings.

⁴PSACs = pre-school-age children; SACs = school-age children.

Characteristics	N (%)
Low (<10% prevalence)	14 (13.9%)
Medium (10–49% prevalence)	65 (64.4%)
High (≥50% prevalence)	22 (21.8%)
Koppen-Geiger climate zone	
Arid, desert, hot	13 (12.9%)
Arid, steppe, hot	6 (5.9%)
Temperate, dry summer, warm summer	1 (1.0%)
Temperate, dry winter, hot summer	3 (3.0%)
Temperate, dry winter, warm summer	4 (4.0%)
Temperate, no dry season, hot summer	10 (9.9%)
Tropical, monsoon	11 (10.9%)
Tropical, rainforest	3 (3.0%)
Tropical, savannah	50 (49.5%)
Water setting	
River	58 (57.4%)
Stream	35 (34.7%)
Lake	24 (23.8%)
Pond	21 (20.8%)
Dam/Reservoir	10 (9.9%)
Canal	10 (9.9%)
Field	3 (3.0%)
Swamp	2 (2.0%)
Number of waterbodies studied	
Single	65 (64.4%)
Multiple ⁵	36 (35.6%)
Nature of waterbody	
Artificial waterbody	19 (18.8%)
Natural waterbody	82 (81.2%)
Flow of waterbody	
Flowing waterbody	46 (45.5%)
Stagnant waterbody	55 (54.5%)
Measurement tool	
Direct water contact observation	4 (4.0%)
Survey	97 (96.0%)
Water contact dimension	
Any water contact	33 (32.7%)
Frequency	27 (26.7%)
Duration	6 (5.9%)
Activity	70 (69.3%)
Study design	
Case-control	4 (4.0%)
Cohort	10 (9.9%)
Cross-sectional	86 (85.1%)
Randomised controlled trial	1 (1.0%)
Study quality	
Low	46 (45.5%)
Moderate	49 (48.5%)

Continued on next page

⁵Multiple different water settings included in the study (e.g. river and lake).

Characteristics	N (%)
High	6 (5.9%)

Ten studies (10%) explicitly mentioned measuring human water contact as one of their study aims. The remaining 91 studies (91%) had a different primary focus but still reported on water contact. Studies included in the systematic review were from 22 different countries across three continents (Africa, Asia, South America; Fig. 2.2a). With 10–13 studies from each country, Brazil, Ethiopia, Nigeria, and Cameroon represented approximately 50% of all included studies. The earliest included study was from 1984, with most studies conducted after 2010 (Fig. 2.2b).

The most frequently studied water settings were as follows: 58 studies (57%) reported on rivers, 35 studies (35%) on streams, 24 studies (24%) on lakes, and 21 studies (21%) on ponds. Thirty-six studies (36%) mentioned multiple types of waterbodies, while 65 studies (65%) mentioned a single type of waterbody.

Studies screened at the full-text stage used four different water contact measurement tools: surveys, direct water contact observation by trained human observers, participant water contact diaries, and GPS loggers. No diary-based study or GPS logger-based study provided required ORs or sufficient information to reconstruct ORs. Of the 101 included studies, 98 studies (97%) used surveys, and four studies (3%) used direct water contact observation. Water contact activities were reported in 70 studies (69%), followed by any water contact, which was reported in 33 studies (33%). Twenty-seven studies (27%) reported on water contact frequency, while six studies (6%) reported on duration. All activity measures came from surveys. Of the four direct observation studies, three recorded whether people had any water contact, two recorded information on duration, and one recorded frequency.

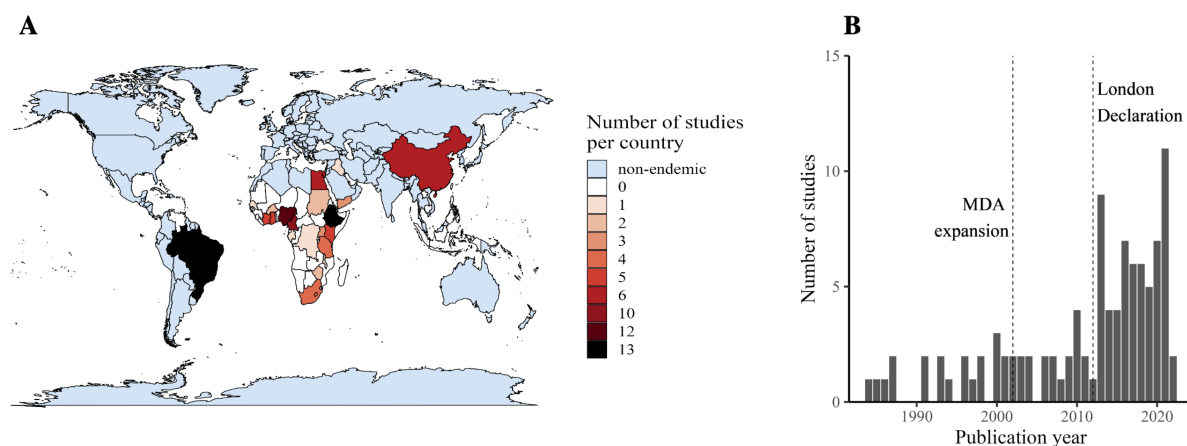


Figure 2.2. Distribution of studies across geography and time

2.3.2 Association of Having Any Water Contact with Infection Status

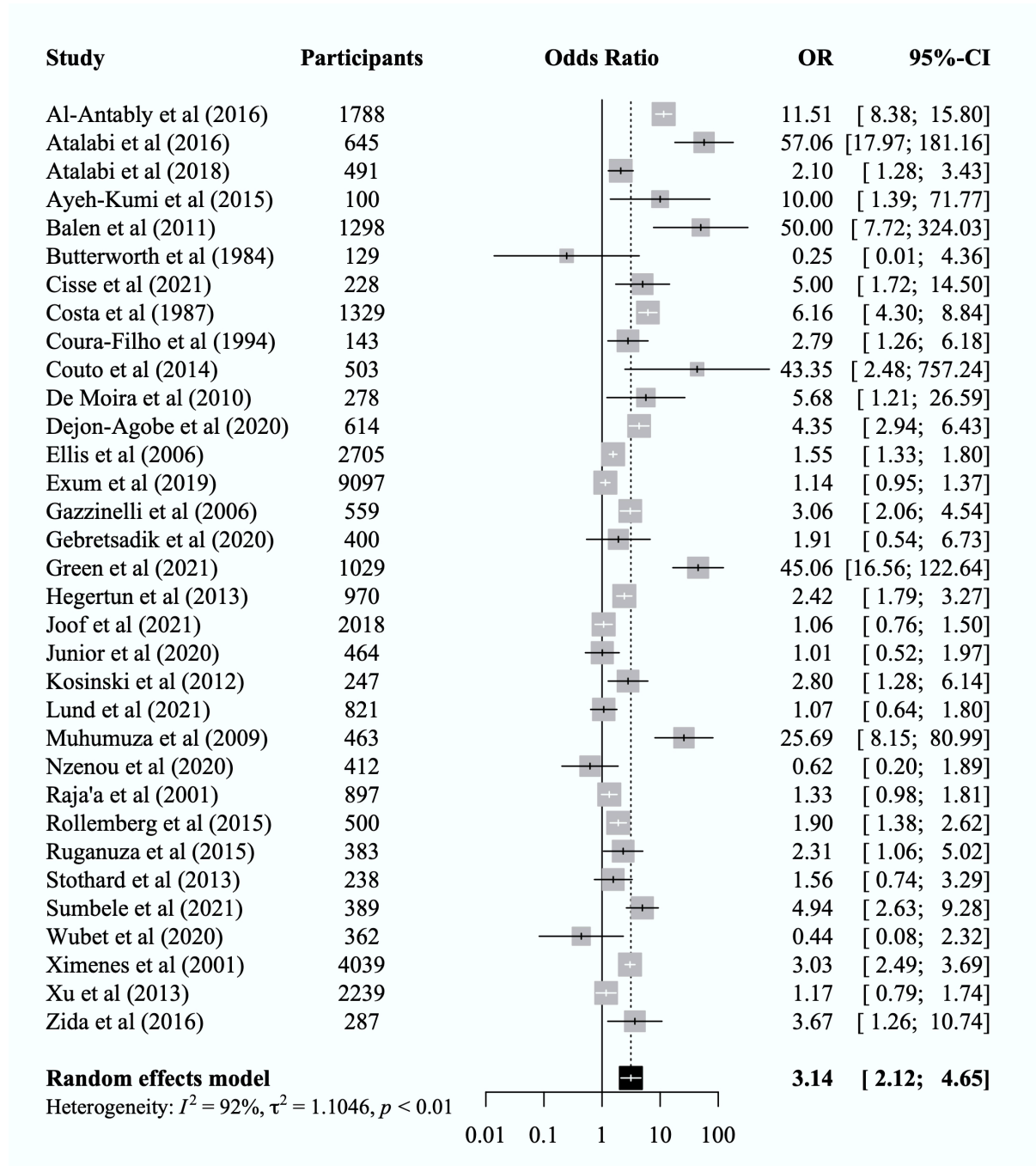


Figure 2.3. Meta-analysis of the association between individual-level water contact and odds of schistosome infection. Pooled effect sizes are based on random-effects models using data from 33 studies representing 36,065 participants. OR = odds ratio. CI = confidence interval. I^2 = I-squared statistic of heterogeneity.

There were 33 studies that reported on associations between any current water contact and schistosome infection status. Included studies represented 36,065 participants, of whom 9839 were infected with a schistosome species. The meta-analysis showed that having any water contact was associated with 3.14 times higher odds of infection (95% CI: 2.12–4.65) when compared to no water contact (Fig. 2.3).

There was no significant difference between studies reporting adjusted effect sizes versus unadjusted effects, but only 9 out of 33 studies provided any adjusted effect sizes (Fig. S2.1). When studies were separated by reinfection versus infection as the outcome, the effect on reinfection was not significant, but the reinfection subgroup included only three studies (Fig. S2.2). When stratified by study quality (grouped into low, medium, and high quality), pooled associations remained significant in all three strata, and no significant differences between groups were observed (Fig. S2.3).

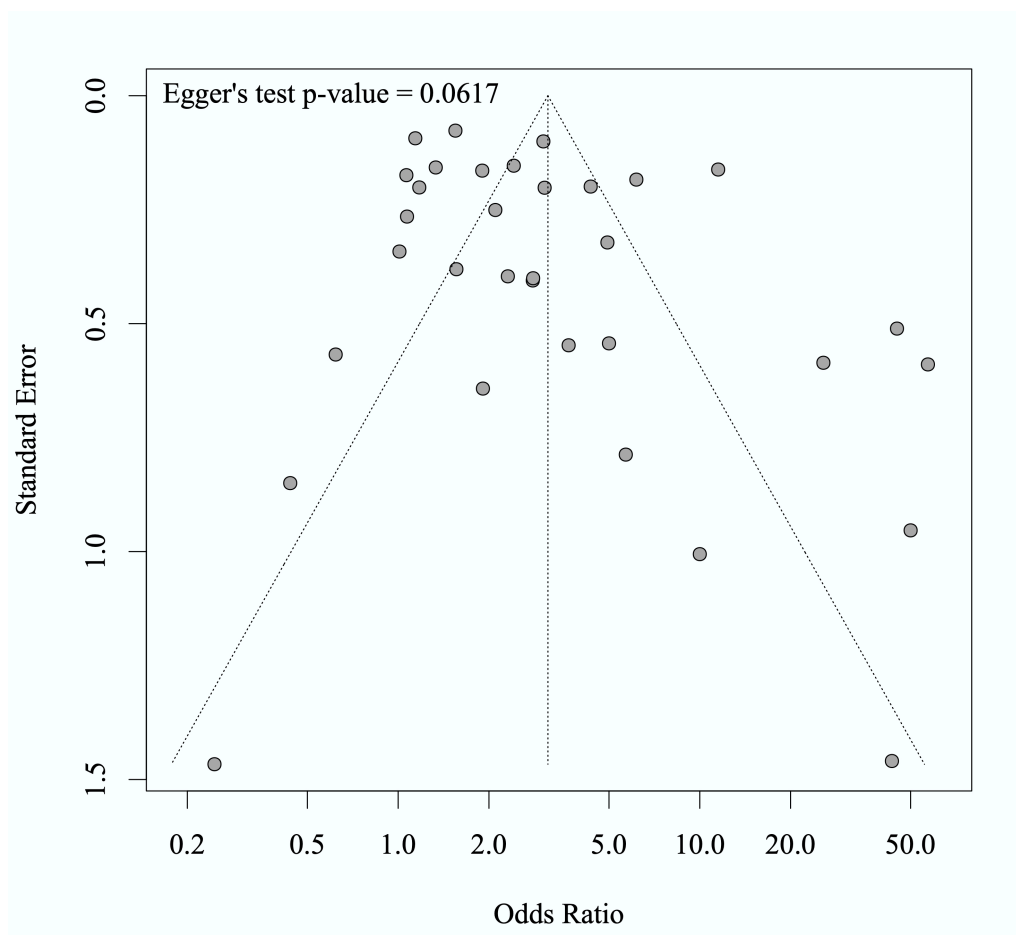


Figure 2.4. Funnel plot for meta-analysis of the association between individual-level water contact and odds of schistosome infection.

No evidence for publication bias was found (Egger's test = 0.06; funnel plot provided in Fig. 2.4). A post-hoc power analysis demonstrated that this meta-analysis had $\geq 99\%$ power to detect the pooled effect of 3.14 that was observed (Fig. S2.4). The meta-analysis was robust to removing any one study at

a time from the analysis (Fig. 2.3). Study quality was rated as high in five studies (15%), moderate in 14 studies (42%), and low in 14 studies (42%) (see Online Table S4 for quality appraisal results [19]).

Subgroup analyses (Fig. 2.5) showed that associations of any water contact with infection were significantly different in studies including only PSACs and SACs compared to studies that included all age groups. Among PSACs and SACs, having water contact was associated with 1.67 times higher infection likelihood, which was lower than the association in studies across all age groups (PSACs, SACs, and adults combined), where water contact conferred 4.24 times greater infection likelihood ($\chi^2 = 7.06$, $df = 1$, $p < 0.01$).

To understand the contribution of schoolchildren to this result, studies were separated by setting (Fig. S2.6), and it was found that water contact was significantly associated with infection in school-based studies (i.e., studies that included schoolchildren) but not in community-based studies (which included both SACs and PSACs). The exclusion of all school-based studies (Fig. S2.7) did not change the finding that water contact was less strongly associated with infection in children compared to studies including children and adults. No overall difference in the effect of water contact on infection between community-based and school-based studies was observed (Fig. 2.5).

Community prevalence, type of water body, and climate zone modified the influence of water contact. In moderate endemicity settings, having water contact was associated with a 3.04 times higher likelihood of infection (95% CI: 1.88–4.89). In high endemicity settings, the respective OR was 4.93 (95% CI: 1.99–12.22). There was no significant association between water contact and odds of infection in low-endemicity settings.

Among the waterbodies studied here, water contact with streams and lakes, but not with rivers, was associated with higher odds of schistosome infection. Reporting contact with streams was associated with 8.16 times higher infection likelihood (95% CI: 3.08–21.61) compared to no water contact. Contact with lakes was associated with 3.20 times higher odds of infection (95% CI: 1.09–9.37).

The subgroup analysis in Fig. 2.5, using the Köppen-Geiger climate zones, showed that among the tropical, arid, and temperate zones represented here, only water contact in tropical savannah climate, which is characterised by a pronounced dry season, was significantly associated with the odds of infection (OR 3.35; 95% CI: 2.22–5.06). This climate zone is found among included studies from sub-Saharan Africa and Brazil. The schistosome species present (*S. haematobium* versus *S. mansoni*) did not significantly modify the odds of infection. Descriptive data on the distribution of schistosome species across different waterbodies among the included studies suggested that *S. haematobium* was more

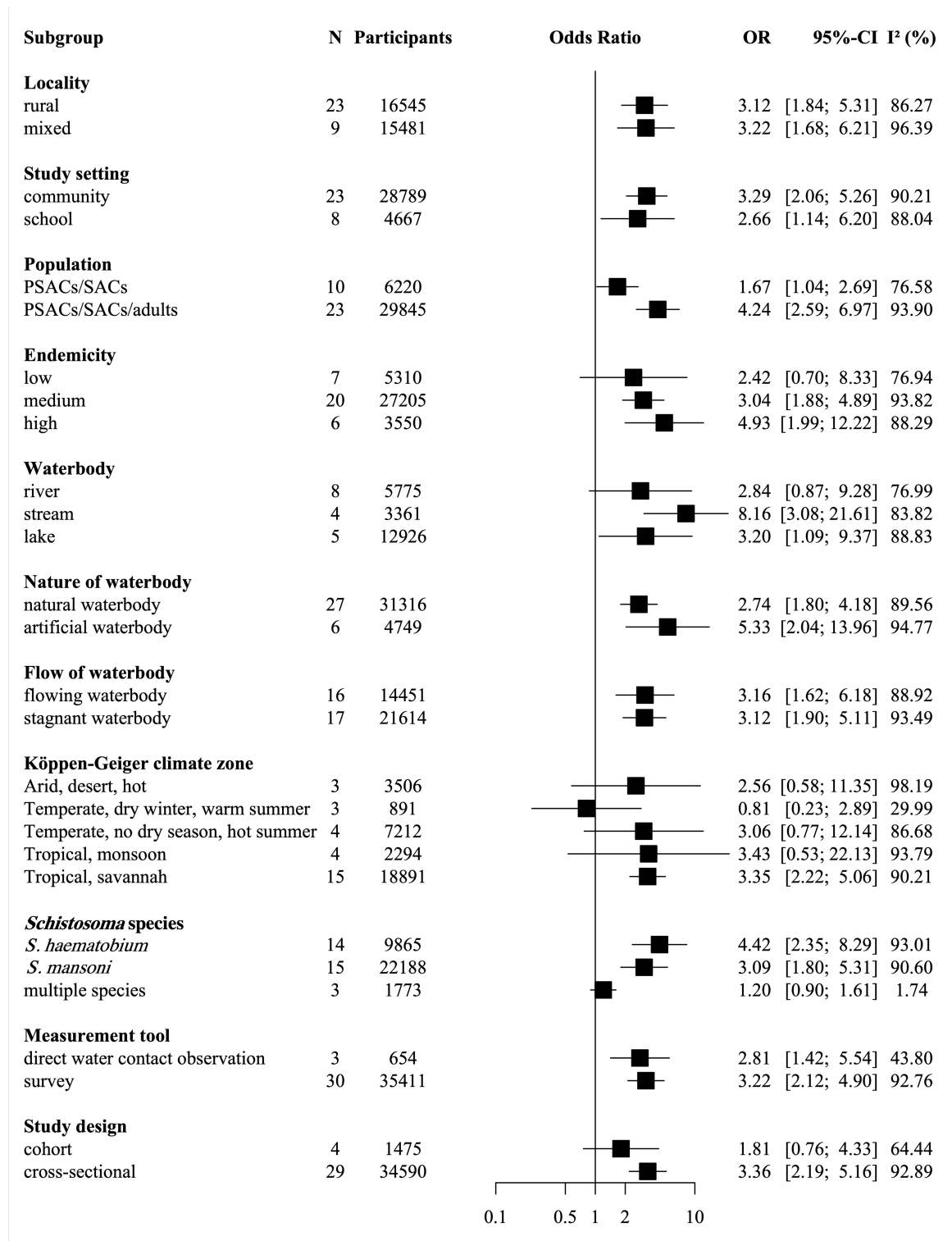


Figure 2.5. Subgroup meta-analysis of the association between individual-level water contact and odds of schistosome infection. Random-effects models using data from 33 studies representing 36,065 participants. N = number of studies. Categories with N < 3 were not reported. OR = odds ratio. CI = confidence interval. I² = I-squared statistic of heterogeneity. Locality = “mixed”: study settings which include urban/peri-urban areas. PSAC = pre-school-age children. SAC = school-age children. Endemicity settings low/medium/high = WHO definitions (low < 10% prevalence, moderate = 10–49% prevalence, high ≥ 50% prevalence). Artificial waterbody = canal, dam/reservoir. Stagnant waterbody = canal, dam/reservoir, lake, pond, field. Köppen-Geiger climate zone = climate zone at the study location.

frequently studied in publications that focused on rivers or streams, whereas *S. mansoni* was mostly the focus of studies that took place along lakes.

Associations of water contact and schistosome infection were significant in cross-sectional studies (OR 3.36; 95% CI: 2.19–5.16) but not in cohort studies (OR 1.81; 95% CI: 0.76–4.33). However, there were only four cohort studies compared to 29 cross-sectional studies. Subgroup analyses also highlighted several study characteristics which did not significantly modify associations between water contact and infection likelihood. Associations remained significant across all types of localities (rural and mixed localities), artificial waterbodies, natural waterbodies, and stagnant and flowing waterbodies. Associations in Fig. 2.5 were also similar across measurement tools (surveys: OR 3.22; 95% CI: 2.12–4.90 and direct water contact observation: OR 2.81; 95% CI: 1.42–5.54). Yet, heterogeneity was substantially lower when water contact was measured through direct observation (I^2 -statistic of 44% versus 93% for surveys).

2.3.3 Association of Water Contact Frequency and Duration with Infection Status

For water contact frequency, 22 studies representing 19,208 participants, of whom 5957 were infected, were meta-analysed (Fig. 2.6). Five studies could not be grouped into these categories as they either had unspecific frequency categories such as ‘sporadic water contact’, reported frequency only on a subgroup, or reported only on infection intensity. Any frequency of water contact (daily, weekly, monthly water contact) was associated with increased odds of infection when compared to no or lower levels of water contact. Although the magnitude of the ORs increased with more frequent water contact, these subgroup differences were not significant ($\chi^2 = 1.69$, $df = 2$, $p = 0.43$). Heterogeneity across subgroups was negatively related to water contact frequency and decreased from daily ($I^2 = 91\%$), weekly ($I^2 = 80\%$),

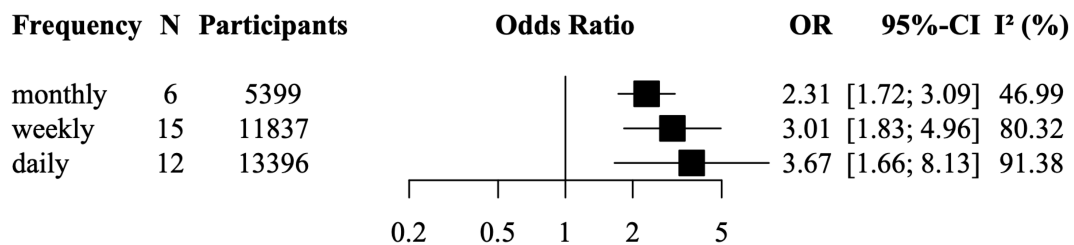


Figure 2.6. Meta-analysis of the association between daily, weekly, and monthly water contact and odds of schistosome infection. Random-effects models using data from 22 studies representing 19,208 participants. N = number of studies. OR = odds ratio. CI = confidence interval. I^2 = I-squared statistic of heterogeneity.

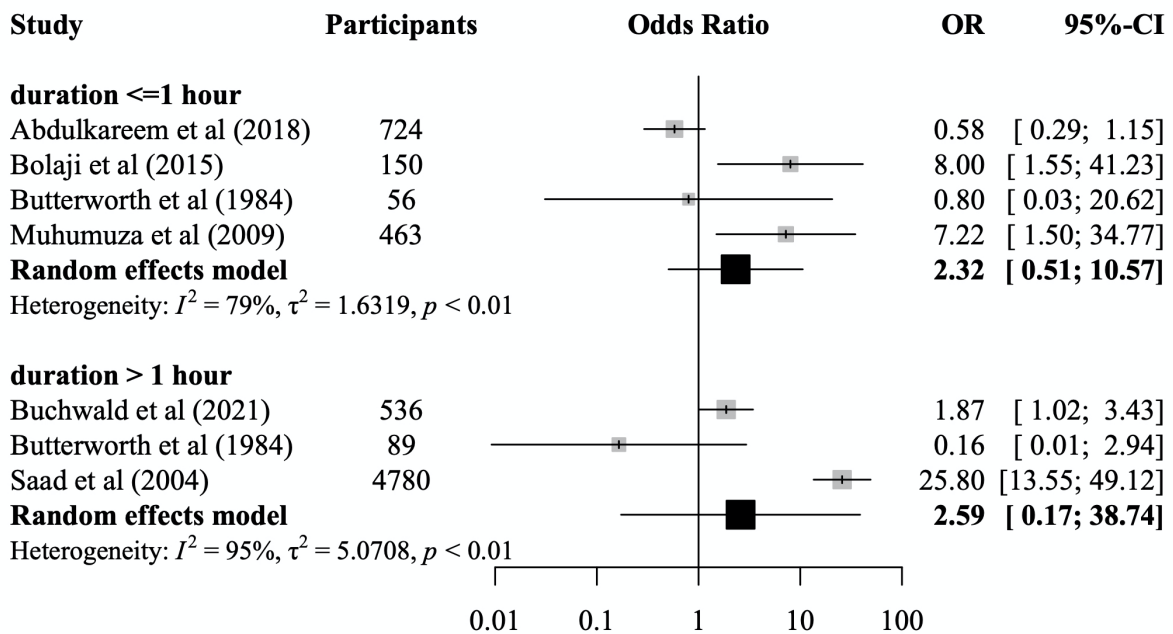


Figure 2.7. Meta-analysis of the association between individual-level water contact duration and odds of schistosome infection. Random-effects models using data from six studies representing 6742 participants. OR = odds ratio. CI = confidence interval. I^2 = I-squared statistic of heterogeneity.

to monthly water contact ($I^2 = 47\%$). Funnel plots and Egger's tests provided no evidence for publication bias (Fig. S2.8). Influence analyses in Fig. S2.9 showed that the analysis was robust to removing any single study from the analysis. Study quality was rated as high in three studies (14%), moderate in ten studies (45%), and low in nine studies (41%).

The meta-analysis of water contact duration was based on six studies representing 6742 participants, of whom 1504 were infected (Fig. 2.7). None of the water contact duration categories (≤ 1 hour, > 1 hour), nor the overall pooled effect size for any duration, was significantly associated with the odds of schistosome infection when compared to no water or lower levels of water contact. Due to the small number of studies in the duration meta-analysis ($N = 6$), no Egger's test or influence analysis was conducted, and the funnel did not provide evidence for asymmetry (Fig. S2.10). Study quality was rated as moderate in four studies (67%) and as low in two studies (33%).

2.3.4 Association of Water Contact Activities with Infection Status

Activities were grouped into distinct activity categories as described in Online Table S3 [19]. Sixty-seven studies representing 102,149 participants, of whom 27,356 were infected, were included in the multilevel meta-analysis. Three studies reported on activity categories that could not be grouped.

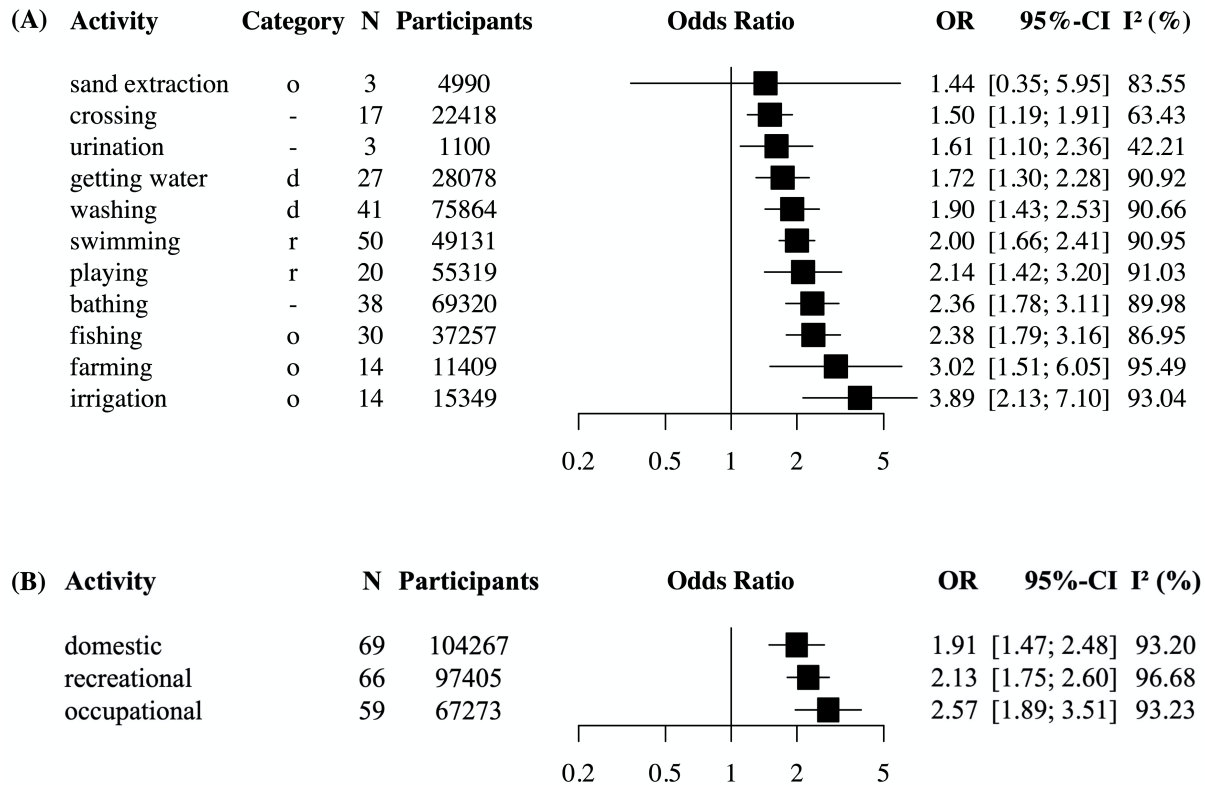


Figure 2.8. Meta-analysis of the associations between water contact activities and odds of schistosome infection. (A) Multilevel random-effects models using data from 67 studies representing 102,149 participants reporting on water contact activities. Washing=washing laundry, animals, clothes, utensils, dishes, or blankets. Farming=farming, agriculture, rice farming. Irrigation = irrigation, watering gardens, or cleaning streams. In the category column, o = occupational, d = domestic, r = recreational, ‘-’ denotes activities that could not be grouped into these categories. (B) Multilevel random-effects models using data from 69 studies with 104,267 participants showing pooled effect sizes according to the water contact categories domestic, recreational, and occupational. N = number of effect sizes. Categories with N < 3 were not reported. OR = odds ratio. CI = confidence interval. I^2 = I-squared statistic of heterogeneity.

In the meta-analysis shown in Fig. 2.8, 10 out of 11 activities were significantly associated with schistosome infection. These activities were crossing waterbodies, open urination, getting water, washing, swimming, playing, bathing, fishing, farming, and irrigation. Sand extraction was the only activity not associated with increased infection risk. Except for crossing and irrigation, the 95% CIs of all activities were overlapping.

In this analysis, three of the included studies (4%) were rated as high quality, 36 (54%) as moderate, and 28 studies (42%) as low quality. When activities were regrouped into broader categories (domestic, recreational, or occupational water contact; see Online Table S3 [19]), occupational activities were most strongly associated with the odds of schistosome infection (OR 2.57; 95% CI 1.89–3.51), followed

by recreational activities (OR 2.13; 95% CI 1.75–2.60) and domestic activities (OR 1.89; 95% CI 1.47–2.48), though their 95% CIs were overlapping.

Heterogeneity was high ($I^2 \geq 90\%$ in most categories), and Egger's tests and funnel plots indicated risk of publication bias (Fig. S2.11). Influence analyses showed that almost all activity categories (except for sand extraction and urination) were robust to removing any single study from the analysis (Figs. S2.12 and S2.13).

2.4 Discussion

Water contact behaviours play a key role in transmission but remain relatively poorly characterised [33]. Since the adoption of World Health Assembly's 54.19 resolution in 2001 [34], MDA using praziquantel has become the cornerstone for schistosomiasis control. However, there is a need to complement MDA with WASH interventions and reductions in water contact to lower schistosomiasis prevalence [26]. To quantify the role of current exposure for transmission, this chapter presents a systematic review and meta-analysis of associations between water contact and schistosome infection, analysing data from 98 studies representing 191,146 participants across three continents. A simple binary measure of having current water contact was strongly associated with infection status across included studies.

2.4.1 Any Water Contact

The association of water contact and infection was large compared to the effect of other risk factors for infection such as WASH or gender. This meta-analysis of 33 studies found that self-reported current water contact was associated with 3.14 times higher odds of infection compared to having no current water contact. By comparison, a meta-analysis by Ayabina et al. [35] found that schistosomiasis prevalence in males was 1.2 times the prevalence in females. The authors hypothesised that gender differences in water contact behaviour may explain their finding. Gender-disaggregated exposure data was rarely reported in the studies, but it is needed to investigate this hypothesis. A meta-analysis by Grimes et al. [36] showed that having no access to safe water was associated with 2.27 times higher odds of infection compared to having access to safe water supplies (estimate transformed from risk ratio to odds ratio for comparability). The relative magnitude of associations of infection status with WASH and with water contact, respectively, is plausible given that water contact is not fully determined by access to WASH. Even when safe water supplies are present, people may continue to have some degree of contact with

open water sources [33].

This chapter shows that the literature on the relationship between water contact and schistosome infection relies mostly on a single tool: cross-sectional surveys. Here, 97 out of 101 studies (96%) were surveys. No diary or GPS logger studies were eligible for inclusion, and only four included studies used direct water contact observation. The dominance of surveys limited the ability to compare how associations varied across tools. Future research should focus on the quantitative analysis of water contact and infection using tools other than surveys to measure water contact. With more varied tools, conclusions could be drawn about the validity of different methods to measure exposure. It was also found that only ten included studies (10%) explicitly stated that their objective was to study the association between water contact and infection. Thus, the body of literature, which mainly focuses on water contact, is small compared to the number of studies that also reported on water contact but had a different primary study focus.

The association of current water contact with schistosome infection suggests that individual-level water contact behaviour might be somewhat stable across the life course. Current water contact cannot explain current infection status. This is because the time from infection to diagnosis is at minimum 4–6 weeks, depending on the diagnostic test used [37, 38]. It has been acknowledged that current exposure is not necessarily representative of the relevant exposure history [39] because water contact varies over time [5, 36]. Nonetheless, cross-sectional studies operate under the assumption that current water contact is an adequate proxy measure of recent, relevant exposure. The same assumption is made in mathematical transmission models. For instance, the stratified-worm-burden model by Gurarie et al. uses cross-sectional water contact and infection data from surveys for model calibration [40]. This thesis chapter reported a robust association between having water contact and infection status, which suggests that current exposure must be at least somewhat representative of past exposure. A similar conclusion has been reached by an observational study in Zimbabwe by Chandiwana and Woolhouse [39]. The authors interpreted associations between schistosome infection intensity and water contact over a two-week observation period as evidence for the persistence of exposure patterns over time. They noted that while there was seasonal variation in contact, individuals with high water contact tended to maintain high contact across seasons. Future studies with multiple timepoints of exposure measurement are needed to more fully investigate to which extent current and past contact are related.

The positive association between having any water contact and infection was significantly attenuated in studies that included only PSACs and SACs compared to studies which included PSACs, SACs,

and adults. The weaker associations of water contact and infection in PSACs and SACs might seem surprising in light of numerous studies that found children to be at highest risk of infection [13, 41]. The attenuated relationship in PSACs and SACs was robust to excluding all school-based studies. This was done to account for the possibility that children in schools may receive more frequent anti-schistosomal treatments. It does not seem that the school-based sampling can explain why the association between water contact and infection was lower in PSACs and SACs compared to all ages combined. As the difference between the group of PSACs and SACs and the group which represents all ages is that the latter includes adults, it is plausible that the addition of adults could explain the greater effect size in the combined group. As immunity intervenes in the association between water contact and infection, it remains possible that the development of acquired immunity represents a confounding variable that explains the attenuated water contact-infection associations in children. However, due to the lack of studies that sampled adults only, the risk of children versus adults could not be directly compared, and further investigation of this hypothesis was not possible as treatment history and immunological data was not available in the included studies.

The lack of studies reporting on age-disaggregated water contact does not allow any conclusions to be made on the shape of the exposure-infection relationship over age. Exposure and acquired immunity are well-established as explanations for the age dependency in schistosome infection, despite poorly established associations on dose-response relationships of water contact and infection likelihood [41, 42]. To better study the contribution of water contact to the age dependency in infection, there is a need to sample adults in addition to SACs [43, 44] and to report separately on their infection outcomes and water contact behaviours. No evidence was found that the likelihood of infection from water contact was any greater for PSACs and SACs compared to other age groups, suggesting that there is no evidence that children were more predisposed to schistosome infection due to water contact compared to adults.

Past studies have found enrolled SACs to be at lower prevalence of infection compared to non-enrolled SACs [45–47]. This finding is not confirmed here as SACs were more likely to be infected when sampled from schools when compared to PSACs and SACs sampled from communities. Limitations of this finding are that this review did not include community-based studies reporting only on SACs and that there were relatively few studies (four school-based and five community-based studies) that allowed for comparisons between SACs sampled in schools versus PSACs and SACs sampled from communities.

2.4.2 Water Contact Frequency and Duration

This chapter found no clear evidence that higher levels of water contact confer significantly higher schistosome infection risk compared to lower levels of water contact, as measured by water contact duration and frequency. It is plausible that relatively low levels of water contact are sufficient for becoming infected. A study by Moses et al. [48] in Uganda noted that water contact duration of as little as five minutes per day was significantly associated with *S. mansoni* infection risk. Animal studies have demonstrated that skin penetration times for *S. mansoni* cercariae are between 2–5 minutes, providing further evidence that low duration water contacts are sufficient for becoming infected [49]. Therefore, water contact intensity may be less relevant for predicting individual-level infection status than knowledge on whether someone had any water contact [50]. The results presented here are consistent with findings from Moses et al. as even monthly water contact (the most infrequent exposure category) was significantly associated with the likelihood of infection. It remains possible that this meta-analysis was unable to account for potential confounders such as the time of day as well as cercarial densities at water sites which mediate the relationship between duration and infection. These data were not typically reported in the included studies. Another possibility is that the meta-analysis has been unable to establish the true relationship between the degree of exposure and infection status due to differential measurement error across exposure categories. The I^2 statistic of heterogeneity indicates that there is almost twice as much variability in the association of infection within the highest water contact frequency category compared to the lowest frequency category. The increase in within-subgroup heterogeneity with higher water contact frequency may indicate a challenge in reliably measuring more frequent water contact. This argument does not extend to duration where the I^2 for the pooled associations was similar between the two duration categories and in fact lower than for any of the other exposure measures. In this chapter, it was not possible to conduct a meta-analysis of water contact frequency and schistosome infection intensity due to a lack of studies reporting on this outcome. However, most research on the dose-response relationship between schistosome infection and degree of exposure has focused on infection intensity [6, 14, 51]. The dose-response relationship between water contact and infection intensity is especially relevant for transmission because heavily infected individuals may disproportionately affect local transmission [52, 53]. Transmission dynamics have been incorporated into mathematical models to guide schistosomiasis control efforts, including those by the WHO [26]. The recently developed stochastic individual-based transmission model SCHISTOX uses an age-adjusted water contact frequency and a cercaria uptake rate to model exposure [54]. Yet, it is currently unclear how exposure should be parametrised in transmission

models when, as observed in this meta-analysis, few studies report on water contact and infection intensity.

2.4.3 Water Contact Activities

Almost any type of water contact activity was linked to increased infection risk despite expected differences in activities regarding contact duration and degree of bodily immersion [6, 15, 55]. The results of this chapter contrast with a previous systematic review (without meta-analysis) by Grimes et al. which considered laundry, bathing and recreational swimming as the activities with most exposure and getting water as unimportant [36]. The meta-analysis reported here did not provide support for a clear ranking of activities, surprisingly so as fishing has been a focus of many studies on schistosome risk [4, 8, 48, 56, 57]. The grouping of activities into high and low risk is implied by the 2006 WHO guidelines which define fishermen, farmers, irrigation workers, and women executing domestic tasks as at-risk adult groups [50]. In this analysis, very few studies reported on schistosome infection likelihood for different at-risk groups. For instance, three studies (3%) reported on fishing as an occupation. At the same time, 29 studies (29%) reported on fishing as an activity but only six of them (21%) adjusted for any covariates, despite the availability of variables such as age and gender, which are well-known risk factors for infection [13, 35, 58]. This chapter hence mostly included crude ORs for the infection likelihood of fishing. When grouping water contact activities into occupational, domestic, and recreational activities, there was no evidence that these categories conferred substantially different risks. To better understand the extent to which various activities are associated with different likelihoods of infection, future research needs to collect more detailed individual-level exposure information, not only on the type of activity, but also on frequency, duration, time of day, cercarial exposure or degree of immersion of that activity. An insufficient number of included studies reported information that would have enabled subgroup meta-analyses of different activities, separated by frequency and duration.

2.4.4 WASH Thresholds and Community Infection Prevalence

This analysis found that having water contact in high prevalence areas ($\geq 50\%$ prevalence) was associated with similar schistosome infection likelihood compared to water contact in medium prevalence areas (10–49%), using the 2022 WHO schistosomiasis control guideline definitions [26]. To which extent community prevalence of schistosomiasis can be explained by water contact behaviour is still unclear. A study in Senegal of an immunologically naïve population found that water contact levels were low

and could not explain the very high community prevalence of schistosomiasis of 75%-100% [6]. By contrast, another study in Kenya found that persistent hotspot villages with poor response to annual MDA treatment differed from lower prevalence villages in their water contact patterns [59]. The hotspot villages had significantly higher proportions of individuals who engaged in bathing, washing clothes or dishes, getting water, playing, and swimming. This chapter reported that infection risk of water contact was similar across high-prevalence and medium-prevalence communities, suggesting that WASH and health education or behaviour change interventions that reduce water contact should not be targeted exclusively at high endemicity settings. In terms of WASH, the results suggest that achieving a critical threshold of (near-)universal community-level coverage may be needed for WASH to successfully control transmission. As almost any water contact activity was associated with infection risk, lowering community-level water contact and contamination may require not just access to safe drinking water, but providing a whole suite of facilities including laundry basins, showering or bathing facilities, toilets, or swimming facilities that minimise all types of open water contact. In this chapter, it was not possible to directly assess how access to safe water or sanitation modified individual-level risk of water contact using subgroup analysis as none of the 33 studies in our main analysis provided effect sizes adjusted by access to safe drinking water or sanitation. Consequently, this meta-analysis was unable to account for WASH or socio-economic factors that may confound associations between water contact and infection. This same limitation of potential socio-economic confounding had been identified in the review by Grimes et al. [36], highlighting the need for studies that combine behavioural, environmental, WASH, and socio-demographic information. For MDA, the implication is that treatment should extend beyond high endemicity settings and needs to reach all groups with regular water contact as the likelihood of (re-)infection conferred by water contact was similar in high and medium endemicity settings.

This meta-analysis found that water contact with streams and lakes was significantly associated with likelihood of infection while contact with rivers was not. It was found that artificial waterbodies (such as reservoirs, dams, and canals) were associated with a similar likelihood of individual-level infection when compared to water contact with natural waterbodies. These results contrast with a meta-analysis by Steinmann et al. [60] which found that prevalence in irrigated areas and dammed landscapes was significantly higher when compared to pre-construction prevalence or prevalence in control areas. These differences could be due to different comparators. Steinmann et al. compare dams and irrigated areas (i.e., settings with favourable environmental conditions for transmission) against control areas with lower environmental risk while our review compared different types of waterbodies against each other, all of

which may be high-risk settings. It is also unclear if differences in infection likelihood across water settings are driven primarily by environmental or behavioural factors. If the latter were the case, one would expect to see no difference in infection likelihood between water settings once water contact behaviour is accounted for. The effect of dam construction on schistosomiasis, for example, could be driven by increased water contact through fishing or irrigation as environmental modification provides new economic opportunities. Such an occupational shift has been observed in Nigeria where fishing became an important water contact activity after the construction of the Ruwan Sanyi dam [56]. More research is needed to understand the aquatic biology and behavioural patterns which shape infection outcomes associated with different waterbodies [61, 62]. It is unclear if the finding of differential risk of different waterbodies (i.e., between rivers, lakes, and streams)—rather than reflecting substantial risk differences in these settings—may be spurious as some waterbodies are more well-studied than others. In this meta-analysis, more than half of all included articles studied streams, one-third studied rivers, and a quarter studied lakes (multiple mentions were possible). By contrast, dams/reservoirs, canals, fields/rice paddies, and swamps provided too few studies for subgroup analyses. Therefore, this analysis had limited ability to contrast risk across the full range of water settings. The finding that only water contact in high and medium endemicity settings, but not low endemicity settings, was associated with increased odds of infection may be due to different likelihoods of parasite acquisition because of variation in environmental hazard. However, the lack of malacological data in most studies precluded an investigation of how the risk of water contact was mediated by the presence of intermediate snail hosts across different waterbodies. Among 101 studies, only 21.8% reported on snail abundance (from malacology) and 12.8% reported on prevalence of infected snails. The number of studies per subgroup was too low to analyse how risk of water contact varied between studies that found snails compared to studies that found no snails.

Seasonality modified associations between having water contact and schistosome infection. Having any water contact was associated with infection status only in study locations classified as tropical savannah climate by the Köppen–Geiger climate zone classification [27]. Tropical savannah climate is characterised by a pronounced dry season either in winter or summer. None of the other climate zones represented in the review had a similarly pronounced dry season. Seasonality has been shown to affect transmission both through changes in human behaviour with higher water contact in the dry season [5] and through variation in environmental conditions (such as water flow, rainfall, and temperature) that affect environmental risk [63, 64]. The findings in this chapter are consistent with previous research that has linked dry-season water contact to elevated infection risk [60]. Moreover, this meta-analysis showed

that within the tropical savannah studies, it was contact with seasonal (as opposed to perennial water sources) which dominated. Among included studies, 50% of waterbodies in tropical savannah climate were streams (a higher proportion than in other climate zones) and only 14% were rivers. Streams are more likely to be seasonal compared to rivers which lends further support to the hypothesis of seasonal water contact as a risk factor.

2.4.5 Limitations

A limitation of this review was the high degree of heterogeneity ($I^2 > 90\%$). The I^2 value was comparable to heterogeneity in two other meta-analyses [35, 36] of observational studies on risk factors for schistosomiasis infection whose I^2 statistics ranged between 82% and 96%. Heterogeneity was lower only in the meta-analysis of associations between water contact duration and infection status where the I^2 was between 65% and 67%. None of the subgroup analyses conducted substantially reduced heterogeneity in the association between schistosome infection and having any water contact, except for the subgroup of direct observation studies which had half the degree of heterogeneity ($I^2 = 43\%$) compared to the subgroup of surveys ($I^2 = 93\%$). Egger's test indicated significant funnel plot asymmetry in the meta-analysis of water contact activities and infection. The source of this asymmetry is not entirely clear; it could reflect reporting bias, inflated effect sizes from smaller studies or true heterogeneity in associations [65].

Water contact measures may not have been entirely comparable across studies due to the absence of standardised survey instruments and varying recall periods and exposure definitions. Only one study mentioned the use of a standardised module to elicit self-reported water contact [33]. The recall periods employed in surveys varied widely between water contact in the past seven days to water contact in the past year and most studies failed to report their recall period. Some of the water contact activity categories may be overlapping and include somewhat heterogeneous measures, as reflected in the high I^2 . There was an insufficient number of studies reporting on livestock farming, even though human–animal water contact is crucial, especially for understanding risk for zoonotic transmission of schistosomiasis [66]. When grouping activities into broader types (domestic, recreational, and occupational), the same participants may be included multiple times if they conducted several domestic activities of the same type, such as farming and fishing. However, multilevel models were used to account for repeated observations across activities within the same study. In addition, some dimensions of water contact that are relevant to infection likelihood were not represented in this review, such as bodily immersion and water contact by

time of day [5, 67].

2.4.6 Conclusion

This chapter provides evidence that having water contact was robustly associated with infection in both children and adults, across different water and endemicity settings, and across measurement tools. The results highlight the key role that human exposure behaviour plays in explaining individual-level infection outcomes. The pooled effect size of the association is sizeable compared to two meta-analyses of other risk factors; gender and lack of safe water access [35, 36]. The findings have implications for public health policies and interventions. Because all age groups, including adults, and people living in medium and high endemicity settings were at infection risk from water contact, there is a need for community-wide MDA and interventions to reduce exposure across all ages. This is in line with the 2022 WHO treatment guideline, which recommends community-based treatment of all people aged two and above in communities with $\geq 10\%$ prevalence [26]. Fishing is often cited as a high-risk occupation, including in the 2022 WHO guidelines, but very few studies report directly on fishing as an occupation. Most evidence on fishing comes from studies reporting on fishing as an activity, but the majority did not adjust for any covariates. Studies on water contact and infection should include occupation as a covariate. There are several ways to improve measurement and reporting of water contact in studies on schistosomiasis. Survey-based studies should clearly state the recall periods and the survey questions which were used to elicit responses. The comparability of various duration and frequency measures across studies could be improved by using “no water contact” as the reference category across analyses of frequency and duration. Similar to the UNICEF and WHO Joint Monitoring Programme definitions for WASH [50], it would be beneficial to create an expert consensus and assemble standardised tools for exposure measurement. Thus far, the literature has focused on the exposure measures of having any water contact and the type of water contact activities. While the number of studies that included measures of human water contact has increased over time, based on the studies included in this review, there has been no clear evolution in the types of study designs or tools used. Self-reported measures of contact and cross-sectional study designs continue to dominate the literature. Direct water contact observation has become somewhat less common over time, but it was never widely adopted as a measurement method. Despite the increased availability of data loggers and smartphones to record GPS locations, this method is still rarely used to obtain water contact measures. Future research needs to ascertain measures of frequency, duration, and immersion for different water contact activities. In addition, studies should

routinely report on schistosome infection intensity as an outcome, not only infection status. To reduce schistosome transmission and develop more effective policies, more studies still are needed to inform accurate exposure measurements for ascertaining the age-dependency of exposure, identifying high-risk groups, and informing transmission modelling.

References

1. Costa MFFL, Magalhaes MHA, Rocha RS, Antunes CMF, and Katz N. Water-contact patterns and socio-economic variables in the epidemiology of schistosomiasis mansoni in an endemic area in Brazil. *Bulletin of the World Health Organization* 1987; 65:57–66
2. Exum NG, Kibira SPS, Ssenyonga R, Nobili J, Shannon AK, Ssempebwa JC, Tukahebwa EM, Radloff S, Schwab KJ, and Makumbi FE. The prevalence of schistosomiasis in Uganda: A nationally representative population estimate to inform control programs and water and sanitation interventions. *PLoS Neglected Tropical Diseases* 2019 Aug; 13:e0007617. doi: 10.1371/journal.pntd.0007617
3. Falcone B, Park S, Wu HW, Leenstra T, Jiz MA, Jarilla B, McGarvey ST, Kurtis JD, and Friedman JF. Comparison of self-reported and observed water contact measures in *Schistosoma japonicum*-endemic villages in Leyte, The Philippines. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2021 Oct :149. doi: 10.1093/trstmh/traab149
4. Seto EYW, Sousa-Figueiredo JC, Betson M, Byalero C, Kabatereine NB, and Stothard JR. 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. *Geospatial health* 2012 Nov; 7:1. doi: 10.4081/gh.2012.99
5. Fulford AJC, Ouma JH, Kariuki HC, Thiongo FW, Klumpp R, Kloos H, Sturrock RF, and Butterworth AE. Water contact observations in Kenyan communities endemic for schistosomiasis: methodology and patterns of behaviour. *Parasitology* 1996 Sep; 113:223–41. doi: 10.1017/S0031182000082007
6. Sow S, Vlas SJ de, Stelma F, Vereecken K, Gryseels B, and Polman K. The contribution of water contact behavior to the high *Schistosoma mansoni* Infection rates observed in the Senegal River Basin. *BMC Infectious Diseases* 2011 Dec; 11:198. doi: 10.1186/1471-2334-11-198
7. Jordan P. Epidemiology and control of schistosomiasis. *British Medical Bulletin* 1972 Jan; 28:55–9. doi: 10.1093/oxfordjournals.bmb.a070894
8. Kloos H, Fulford A, Butterworth A, Sturrock R, Ouma J, Kariuki H, Thiongo F, Dalton P, and Klumpp R. Spatial patterns of human water contact and *Schistosoma mansoni* transmission and infection in four rural areas in Machakos District, Kenya. *Social Science & Medicine* 1997 Apr; 44:949–68. doi: 10.1016/S0277-9536(96)00218-3
9. Chandiwan SK. Community water-contact patterns and the transmission of *Schistosoma haematobium* in the highveld region of Zimbabwe. *Social Science & Medicine* 1987 Jan; 25:495–505. doi: 10.1016/0277-9536(87)90173-0
10. Paz-Soldan VA, Reiner RC, Morrison AC, Stoddard ST, Kitron U, Scott TW, Elder JP, Halsey ES, Kochel TJ, Astete H, and Vazquez-Prokopec GM. Strengths and Weaknesses of Global Positioning System (GPS) Data-Loggers and Semi-structured Interviews for Capturing Fine-scale Human Mobility: Findings from Iquitos, Peru. *PLoS Neglected Tropical Diseases* 2014 Jun; 8:e2888. doi: 10.1371/journal.pntd.0002888
11. Macklin G, Stanton MC, Tchuem-Tchuente LA, and Stothard JR. A pilot study using wearable global positioning system data loggers to compare water contact levels: *Schistosoma haematobium* infection in pre-school-age children (PSAC) and their mothers at Barombi Kotto, Cameroon. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2018 Jul; 112:361–5. doi: 10.1093/trstmh/try059
12. Woolhouse MEJ, Taylor P, Matanhire D, and Chandiwan SK. Acquired immunity and epidemiology of *Schistosoma haematobium*. *Nature* 1991 Jun; 351:757–9. doi: 10.1038/351757a0
13. Gryseels B. Human resistance to *Schistosoma* infections: Age or experience? *Parasitology Today* 1994 Jan; 10:380–4. doi: 10.1016/0169-4758(94)90226-7

14. Butterworth AE, Dalton PR, and Dunne DW. Immunity after treatment of human schistosomiasis mansoni. I. Study design, pretreatment observations and the results of treatment. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1984; 78:108–23. doi: 10.1016/0035-9203(84)90190-1
15. Wilkins H, Blumenthal U, Hagan P, Hayes R, and Tulloch S. Resistance to reinfection after treatment of urinary schistosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1987 Jan; 81:29–35. doi: 10.1016/0035-9203(87)90273-2
16. Kabatereine NB, Vennervald BJ, Ouma JH, Kemijumbi J, Butterworth AE, Dunne DW, and Fulford AJC. Adult resistance to schistosomiasis mansoni: age-dependence of reinfection remains constant in communities with diverse exposure patterns. *Parasitology* 1999 Jan; 118:101–5. doi: 10.1017/S0031182098003576
17. Reitzug F and Chami GF. Associations of water contact duration and frequency with *Schistosoma* infection risk in different measurement tools: A systematic review and meta-analysis. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42022333680
18. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, and Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021 Mar; 372:n71. doi: 10.1136/bmj.n71
19. Reitzug F, Ledien J, and Chami GF. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2023 Jun; 17:e0011377. doi: 10.1371/journal.pntd.0011377
20. Morgan RL, Whaley P, Thayer KA, and Schünemann HJ. Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment International* 2018 Dec; 121:1027–31. doi: 10.1016/j.envint.2018.07.015
21. Katz N, Chaves A, and Pellegrino J. A simple device for quantitative stool thick-smear technique in Schistosomiasis mansoni. *Revista Do Instituto De Medicina Tropical De Sao Paulo* 1972 Dec; 14:397–400
22. Mott KE, Baltes R, Bambagha J, and Baldassini B. Field studies of a reusable polyamide filter for detection of *Schistosoma haematobium* eggs by urine filtration. *Tropenmedizin Und Parasitologie* 1982 Dec; 33:227–8
23. Deelder AM, Van Dam GJ, Kornelis D, Fillié YE, and Van Zeyl RJM. *Schistosoma*: analysis of monoclonal antibodies reactive with the circulating antigens CAA and CCA. *Parasitology* 1996 Jan; 112:21–35. doi: 10.1017/S0031182000065045
24. Dam GJ van, Wichers JH, Ferreira TMF, Ghati D, Amerongen A van, and Deelder AM. Diagnosis of Schistosomiasis by Reagent Strip Test for Detection of Circulating Cathodic Antigen. *Journal of Clinical Microbiology* 2004 Dec; 42:5458–61. doi: 10.1128/JCM.42.12.5458-5461.2004
25. Pontes LA, Dias-Neto E, and Rabello A. Detection by polymerase chain reaction of *Schistosoma mansoni* DNA in human serum and feces. *The American journal of tropical medicine and hygiene* 2002; 66:157–62
26. World Health Organization. WHO guideline on control and elimination of human schistosomiasis. Geneva: World Health Organization, 2022
27. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, and Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* 2018 Dec; 5:180214. doi: 10.1038/sdata.2018.214
28. DerSimonian R and Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986 Sep; 7:177–88. doi: 10.1016/0197-2456(86)90046-2
29. Schwarzer G. meta: An R package for meta-analysis. <https://cran.r-project.org/web/packages/meta/>. 2022
30. Harrer M, Cuijpers P, Furukawa T, and Ebert DD. dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'. <http://dmetar.protectlab.org/>. 2019
31. R Core Team. R: A language and environment for statistical computing. manual. Vienna, Austria, 2022
32. National Heart L and Institute B. Observational Cohort and Cross-Sectional Studies from the National Institutes of Health. <https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools>

33. Joof E, Sanyang A, Camara Y, Sey A, Baldeh I, Jah S, Ceesay S, Sambou S, Sanyang S, Wade C, and Sanneh B. Prevalence and risk factors of schistosomiasis among primary school children in four selected regions of The Gambia. *PLoS Neglected Tropical Diseases* 2021 May; 15. doi: 10.1371/journal.pntd.0009380
34. World Health Assembly, 54. Fifty-fourth World Health Assembly, Geneva, 14-22 May 2001: resolutions and decisions. Tech. rep. Geneva: World Health Organization, 2001
35. Ayabina DV, Clark J, Bayley H, Lamberton PHL, Toor J, and Hollingsworth TD. Gender-related differences in prevalence, intensity and associated risk factors of *Schistosoma* infections in Africa: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2021 Nov; 15:e0009083. doi: 10.1371/journal.pntd.0009083
36. Grimes JET, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The Relationship between Water, Sanitation and Schistosomiasis: A Systematic Review and Meta-analysis. *PLoS Neglected Tropical Diseases* 2014 Dec; 8:e3296. doi: 10.1371/journal.pntd.0003296
37. Grootveld R van, Dam GJ van, Dood C de, Vries JJC de, Visser LG, Corstjens PLAM, and Lieshout L van. Improved diagnosis of active *Schistosoma* infection in travellers and migrants using the ultra-sensitive in-house lateral flow test for detection of circulating anodic antigen (CAA) in serum. *European Journal of Clinical Microbiology & Infectious Diseases* 2018 Jul; 37:1709–16. doi: 10.1007/s10096-018-3303-x
38. Hinz R, Schwarz NG, Hahn A, and Frickmann H. Serological approaches for the diagnosis of schistosomiasis – A review. *Molecular and Cellular Probes* 2017 Feb; 31:2–21. doi: 10.1016/j.mcp.2016.12.003
39. Chandiwana SK and Woolhouse ME. Heterogeneities in water contact patterns and the epidemiology of *Schistosoma haematobium*. *Parasitology* 1991 Dec; 103 Pt 3:363–70. doi: 10.1017/s0031182000059874
40. Gurarie D, King CH, Yoon N, and Li E. Refined stratified-worm-burden models that incorporate specific biological features of human and snail hosts provide better estimates of *Schistosoma* diagnosis, transmission, and control. *Parasites & Vectors* 2016 Dec; 9:428. doi: 10.1186/s13071-016-1681-4
41. Hotez PJ, Alvarado M, Basáñez MG, Bolliger I, Bourne R, Boussinesq M, Brooker SJ, Brown AS, Buckle G, Budke CM, Carabin H, Coffeng LE, Fèvre EM, Fürst T, Halasa YA, Jasrasaria R, Johns NE, Keiser J, King CH, Lozano R, Murdoch ME, O’Hanlon S, Pion SDS, Pullan RL, Ramaiah KD, Roberts T, Shepard DS, Smith JL, Stolk WA, Undurraga EA, Utzinger J, Wang M, Murray CJL, and Naghavi M. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. *PLoS Neglected Tropical Diseases* 2014 Jul; 8:e2865. doi: 10.1371/journal.pntd.0002865
42. Pinot de Moira A, Fulford AJC, Kabatereine NB, Ouma JH, Booth M, and Dunne DW. Analysis of Complex Patterns of Human Exposure and Immunity to *Schistosomiasis mansoni*: The Influence of Age, Sex, Ethnicity and IgE. *PLoS Neglected Tropical Diseases* 2010 Sep; 4:e820. doi: 10.1371/journal.pntd.0000820
43. Toor J, Turner HC, Truscott JE, Werkman M, Phillips AE, Alsallaq R, Medley GF, King CH, and Anderson RM. The design of schistosomiasis monitoring and evaluation programmes: The importance of collecting adult data to inform treatment strategies for *Schistosoma mansoni*. *PLoS Neglected Tropical Diseases* 2018 Oct; 12:e0006717. doi: 10.1371/journal.pntd.0006717
44. Toor J, Hamley JJ, Fronterre C, Castaño MS, Chapman LA, Coffeng LE, Giardina F, Lietman TM, Michael E, and Pinsent A. Strengthening data collection for neglected tropical diseases: What data are needed for models to better inform tailored intervention programmes? *PLoS Neglected Tropical Diseases* 2021; 15:e0009351
45. Maseko T, Mkhonta N, Masuku S, Dlamini S, and Fan C. Schistosomiasis knowledge, attitude, practices, and associated factors among primary school children in the Siphofaneni area in the Lowveld of Swaziland. *Journal of Microbiology, Immunology and Infection* 2018 Feb; 51:103–9. doi: 10.1016/j.jmii.2015.12.003
46. Husein M, Talaat M, El-Sayed M, El-Badawi A, and Evans D. Who misses out with school-based health programmes? A study of schistosomiasis control in Egypt. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1996 Jul; 90:362–5. doi: 10.1016/S0035-9203(96)90506-4
47. Fenwick A, Webster JP, Bosque-Oliva E, Blair L, Fleming FM, Zhang Y, Garba A, Stothard JR, Gabrielli AF, Clements ACA, Kabatereine NB, Toure S, Demele R, Nyandindi U, Mwansa J, and Koukounari A. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002–2008. *Parasitology* 2009 Nov; 136:1719–30. doi: 10.1017/S0031182009990400
48. Moses A, Adriko M, Kibwika B, Tukahebwa EM, Faust CL, and Lamberton PHL. Residence Time, Water Contact, and Age-driven *Schistosoma mansoni* Infection in Hotspot Communities in Uganda. *The American Journal of Tropical Medicine and Hygiene* 2021 Dec; 105:1772–81. doi: 10.4269/ajtmh.21-0391

49. Stirewalt MA and Hackey JR. Penetration of Host Skin by Cercariae of *Schistosoma mansoni*. I. Observed Entry into Skin of Mouse, Hamster, Rat, Monkey and Man. *The Journal of Parasitology* 1956 Dec; 42:565–80
50. World Health Organization. Preventive chemotherapy in human helminthiasis. Coordinated use of anthelmintic drugs in control interventions: a manual for health professionals and programme managers. Geneva: World Health Organization, 2006
51. De Moira AP, Fulford AJC, Kabatereine NB, Kazibwe F, Ouma JH, Dunne DW, and Booth M. Microgeographical and tribal variations in water contact and *Schistosoma mansoni* exposure within a Ugandan fishing community: Water contact and *Schistosoma mansoni* in Uganda. *Tropical Medicine & International Health* 2007 Jun; 12:724–35. doi: 10.1111/j.1365-3156.2007.01842.x
52. Amoah AS, Hoekstra PT, Casacuberta-Partal M, Coffeng LE, Corstjens PLAM, Greco B, Lieshout L van, Lim MD, Markwalter CF, Odiere MR, Reinhard-Rupp J, Roestenberg M, Stothard R, Tchuem Tchuente LA, Vlas SJ de, and Dam GJ van. Sensitive diagnostic tools and targeted drug administration strategies are needed to eliminate schistosomiasis. *The Lancet Infectious Diseases* 2020 Jul; 20:e165–e172. doi: 10.1016/S1473-3099(20)30254-1
53. Vercruysse J, Shaw DJ, and De Bont J. Index of potential contamination for schistosomiasis. *Trends in Parasitology* 2001 Jun; 17:256–61. doi: 10.1016/S1471-4922(01)01937-7
54. Graham M, Ayabina D, Lucas TC, Collyer BS, Medley GF, Hollingsworth TD, and Toor J. SCHISTOX: An individual based model for the epidemiology and control of schistosomiasis. *Infectious Disease Modelling* 2021; 6:438–47. doi: 10.1016/j.idm.2021.01.010
55. Lund AJ, Sam MM, Sy AB, Sow OW, Ali S, Sokolow SH, Berekenyei Merrell S, Bruce J, Jouanard N, Senghor S, Riveau G, Lopez-Carr D, and De Leo GA. Unavoidable Risks: Local Perspectives on Water Contact Behavior and Implications for Schistosomiasis Control in an Agricultural Region of Northern Senegal. *The American Journal of Tropical Medicine and Hygiene* 2019 Oct; 101:837–47. doi: 10.4269/ajtmh.19-0099
56. Tayo MA, Pugh RN, and Bradley AK. Malumfashi Endemic Diseases Research Project, XI. Water-contact activities in the schistosomiasis study area. *Annals of Tropical Medicine and Parasitology* 1980 Jun; 74:347–54. doi: 10.1080/00034983.1980.11687351
57. Lamberti O, Kabatereine NB, Tukahebwa EM, and Chami GF. *Schistosoma mansoni* infection risk for school-aged children clusters within households and is modified by distance to freshwater bodies. *PLoS ONE* 2021 Nov; 16:e0258915. doi: 10.1371/journal.pone.0258915
58. Scott J, Diakhate M, Vereecken K, Fall A, Diop M, Ly A, De Clercq D, Vlas S de, Berkvens D, Kestens L, and Gryseels B. Human water contacts patterns in *Schistosoma mansoni* epidemic foci in northern Senegal change according to age, sex and place of residence, but are not related to intensity of infection. *Tropical Medicine & International Health* 2003 Feb; 8:100–8. doi: 10.1046/j.1365-3156.2003.00993.x
59. Musuva R, Odiere M, Mwinzi P, Omondi I, Rawago F, Matendechero S, Kittur N, Campbell C, and Colley D. Unprotected water sources and low latrine coverage are contributing factors to persistent hotspots for schistosomiasis in western Kenya. *PLoS ONE* 2021 Sep; 16. doi: 10.1371/journal.pone.0253115
60. Steinmann P, Keiser J, Bos R, Tanner M, and Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases* 2006 Jul; 6:411–25. doi: 10.1016/S1473-3099(06)70521-7
61. Lund AJ, Sokolow SH, Jones IJ, Wood CL, Ali S, Chamberlin A, Sy AB, Sam MM, Jouanard N, Schacht AM, Senghor S, Fall A, Ndione R, Riveau G, De Leo GA, and Lopez-Carr D. Exposure, hazard, and vulnerability and their contribution to *Schistosoma haematobium* re-infection in northern Senegal. *The Lancet Planetary Health* 2021 Apr; 5:S10. doi: 10.1016/S2542-5196(21)00094-2
62. Jones IJ, Sokolow SH, Chamberlin AJ, Lund AJ, Jouanard N, Bandagny L, Ndione R, Senghor S, Schacht AM, Riveau G, Hopkins SR, Rohr JR, Remais JV, Lafferty KD, Kuris AM, Wood CL, and De Leo G. Schistosome infection in Senegal is associated with different spatial extents of risk and ecological drivers for *Schistosoma haematobium* and *S. mansoni*. *PLoS Neglected Tropical Diseases* 2021 Sep; 15:e0009712. doi: 10.1371/journal.pntd.0009712
63. Walz Y, Wegmann M, Dech S, Raso G, and Utzinger J. Risk profiling of schistosomiasis using remote sensing: approaches, challenges and outlook. *Parasites & Vectors* 2015 Dec; 8:163. doi: 10.1186/s13071-015-0732-6

64. Stensgaard AS, Vounatsou P, Sengupta ME, and Utzinger J. Schistosomes, snails and climate change: Current trends and future expectations. *Acta Tropica* 2019 Feb; 190:257–68. doi: 10.1016/j.actatropica.2018.09.013
65. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, Carpenter J, Rucker G, Harbord RM, Schmid CH, Tetzlaff J, Deeks JJ, Peters J, Macaskill P, Schwarzer G, Duval S, Altman DG, Moher D, and Higgins JPT. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011 Jul; 343:d4002–d4002. doi: 10.1136/bmj.d4002
66. Léger E, Borlase A, Fall CB, Diouf ND, Diop SD, Yasenev L, Catalano S, Thiam CT, Ndiaye A, Emery A, Morrell A, Rabone M, Ndao M, Faye B, Rollinson D, Rudge JW, Sène M, and Webster JP. Prevalence and distribution of schistosomiasis in human, livestock, and snail populations in northern Senegal: A One Health epidemiological study of a multi-host system. *The Lancet Planetary Health* 2020 Aug; 4:e330–e342. doi: 10.1016/S2542-5196(20)30129-7
67. Farooq M and Mallah MB. The behavioural pattern of social and religious water-contact activities in the Egypt-49 bilharziasis project area. *Bulletin of the World Health Organization* 1966; 35:377–87

2.5 Supplementary Information

Water contact and odds of schistosome infection – subgroup analysis of adjusted versus unadjusted effect sizes

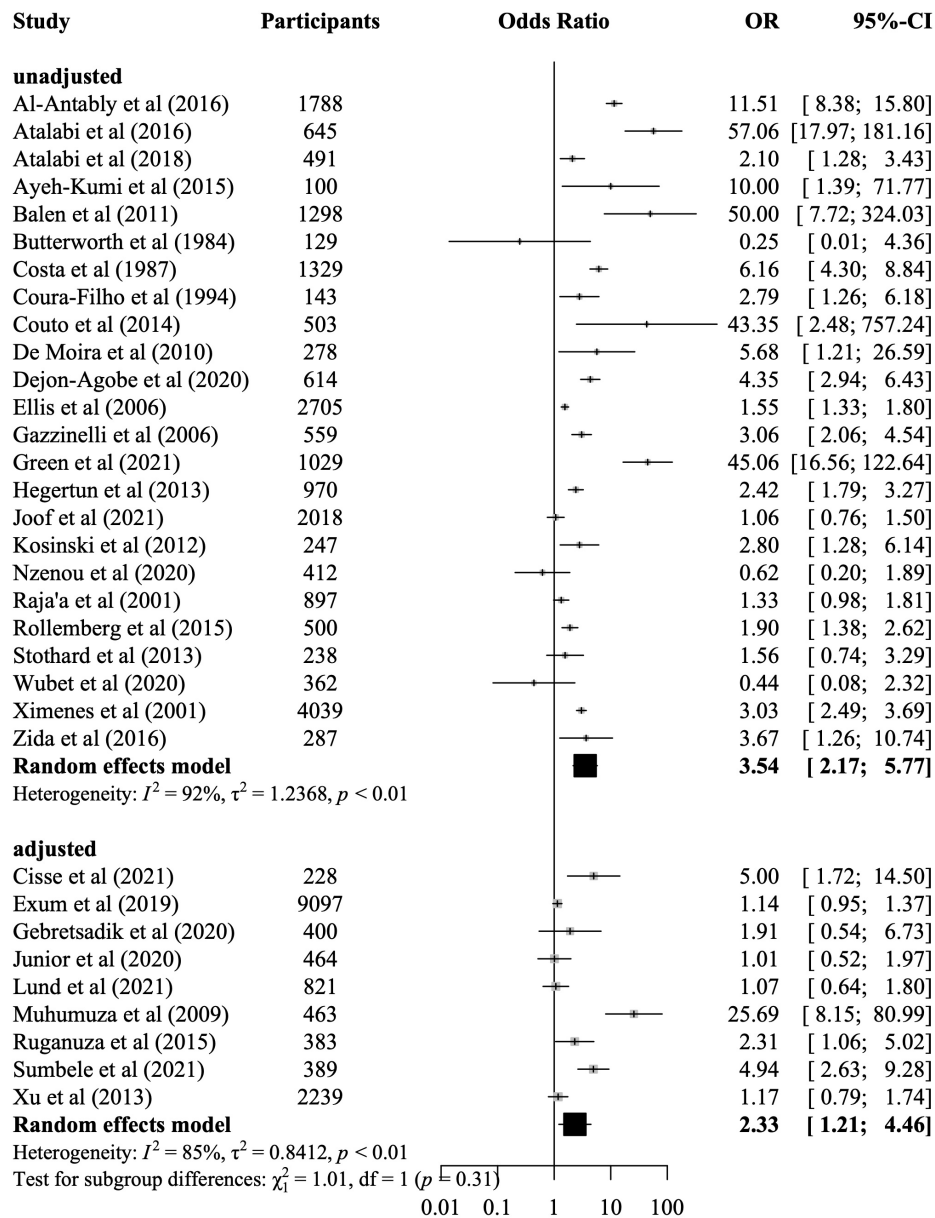


Figure S2.1. Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis of adjusted versus unadjusted effect sizes.

**Water contact and odds of schistosome infection – subgroup analysis
of infection versus reinfection as the outcome**

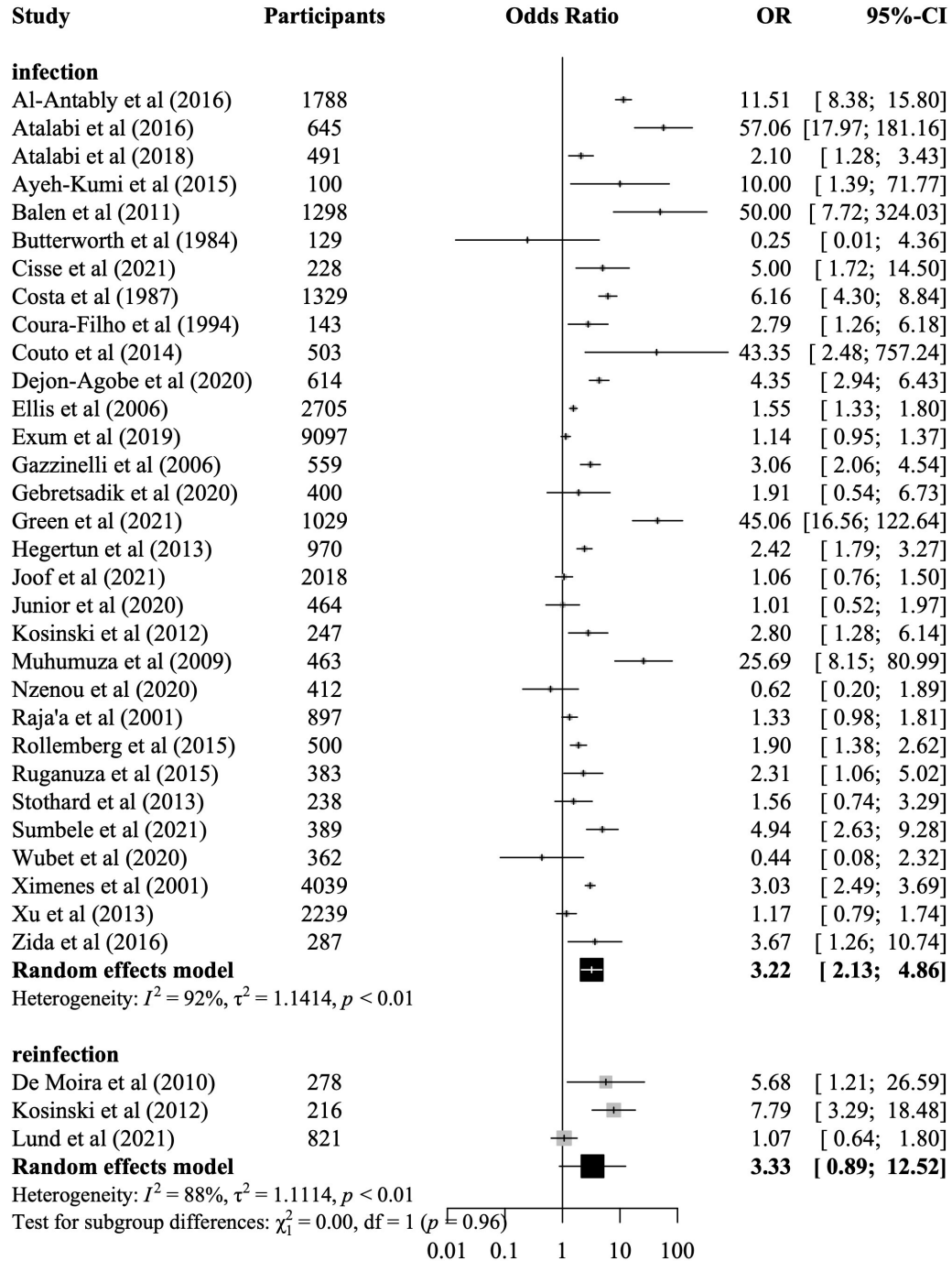


Figure S2.2. Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis of infection/reinfection.

**Water contact and odds of schistosome infection – subgroup analysis
by study quality (low, moderate, high)**

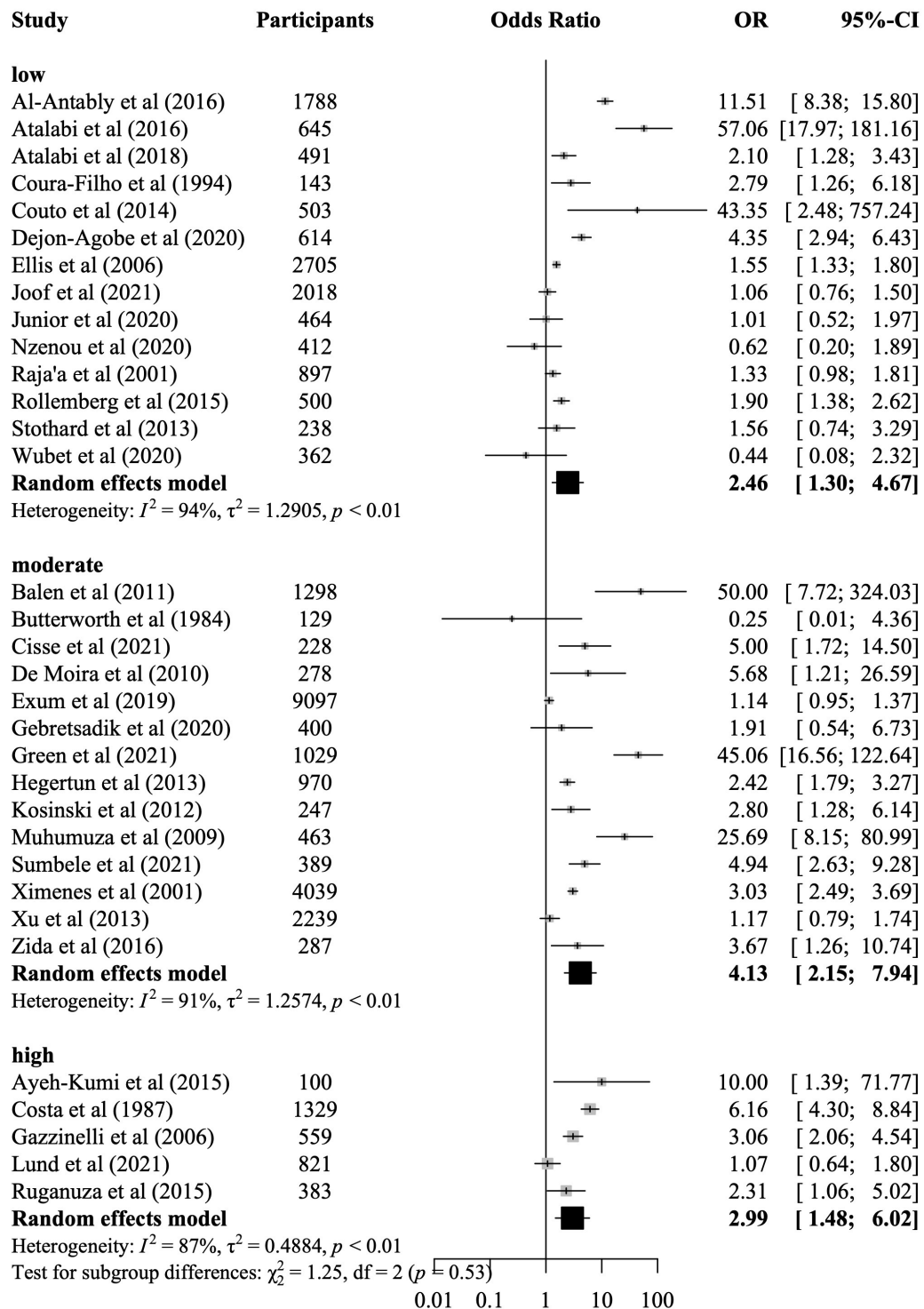


Figure S2.3. Meta-analysis of the association between individual-level water contact and odds of schistosome infection – subgroup analysis by study quality (low, moderate, high).

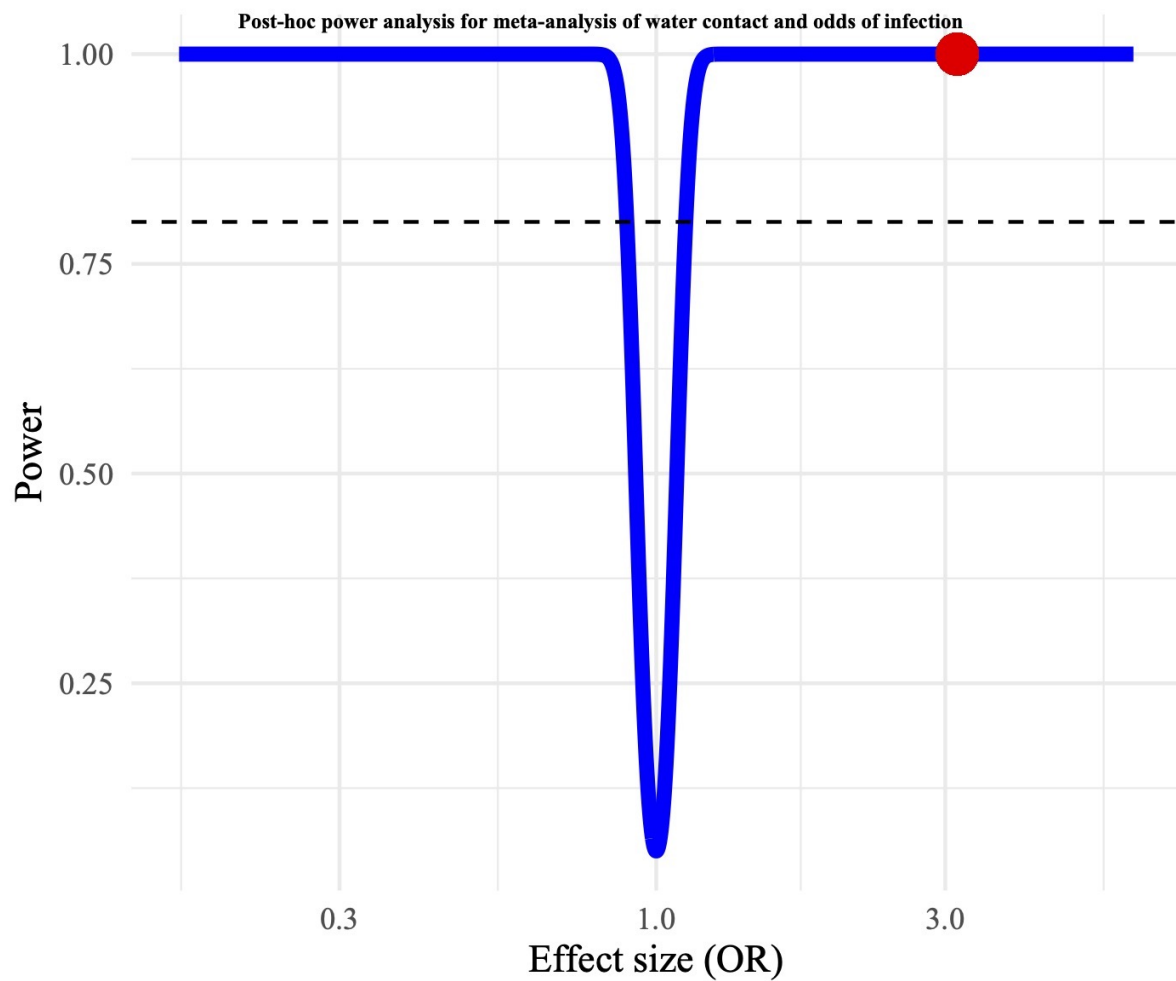


Figure S2.4. Post-hoc power calculation to detect an OR of 3.14 with 31 studies and a median study size of 503 participants >99% (parameters set to values from the meta-analysis).

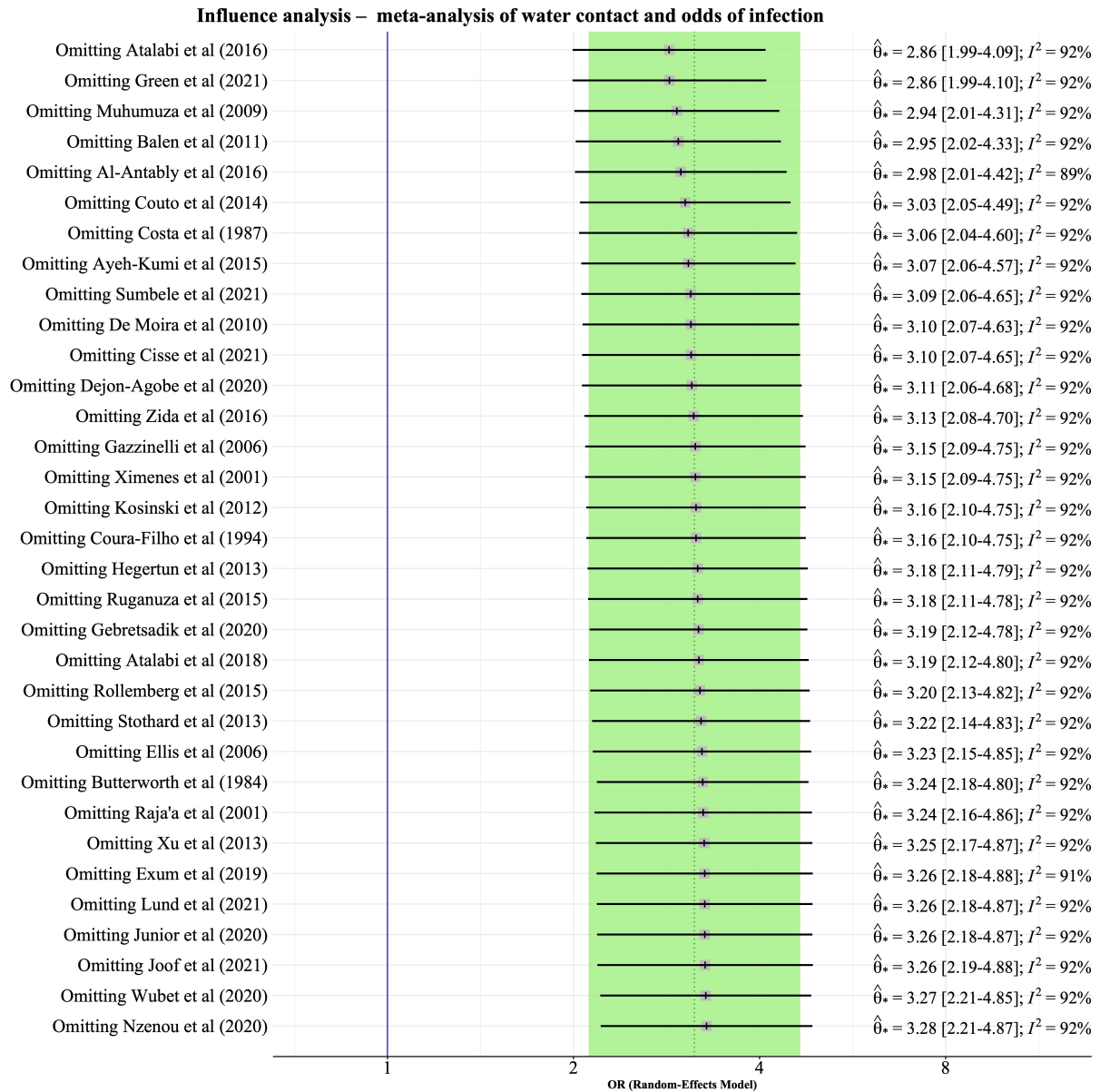


Figure S2.5. Assessing the influence analysis of any single study in the meta-analysis of associations between individual-level water contact and odds of schistosome infection by removing one study at a time.

Subgroup meta-analysis by study setting of the association between individual-level water contact and odds of schistosome infection in children

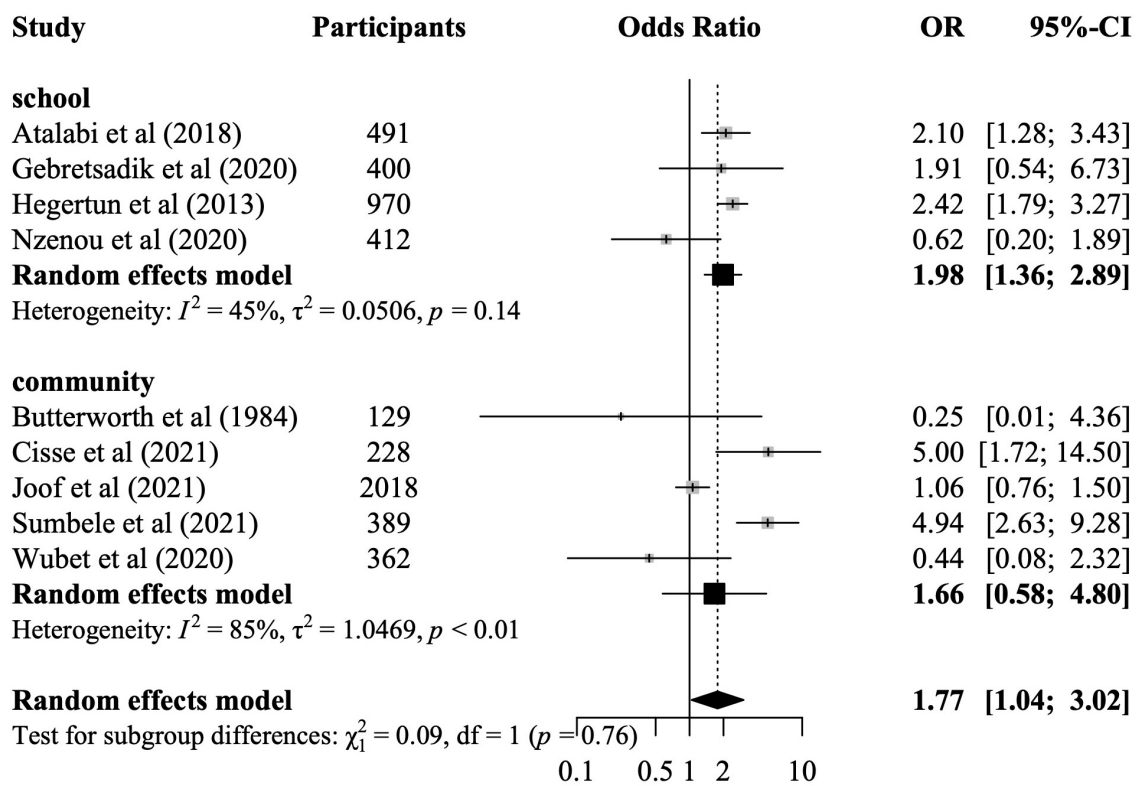


Figure S2.6. Subgroup meta-analysis by study setting of the association between individual-level water contact and odds of schistosome infection in children.

Subgroup meta-analysis of the association between individual-level water contact and odds of schistosome infection in community-based studies

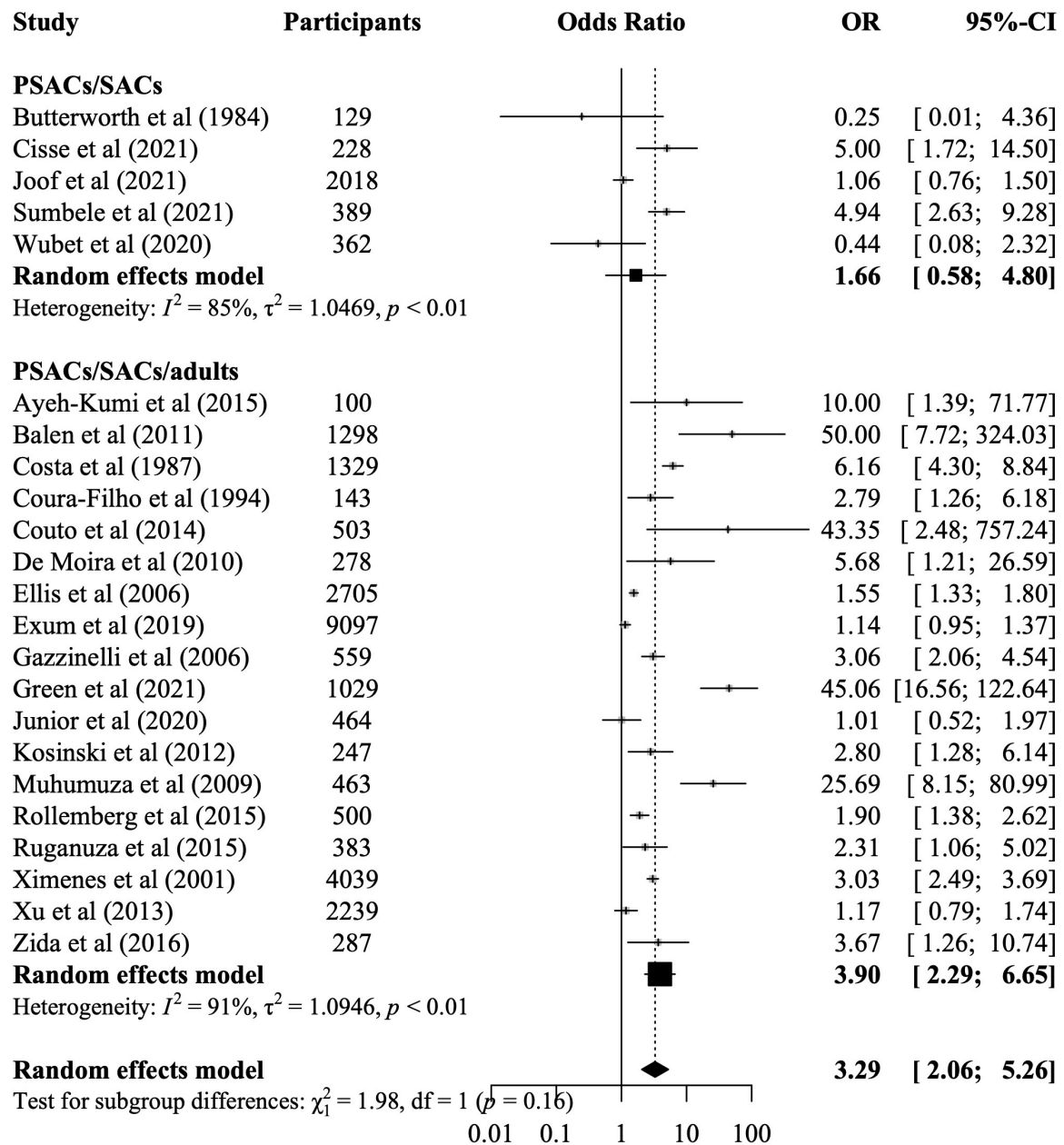


Figure S2.7. Subgroup meta-analysis of the association between individual-level water contact and odds of schistosome infection in community-based studies.

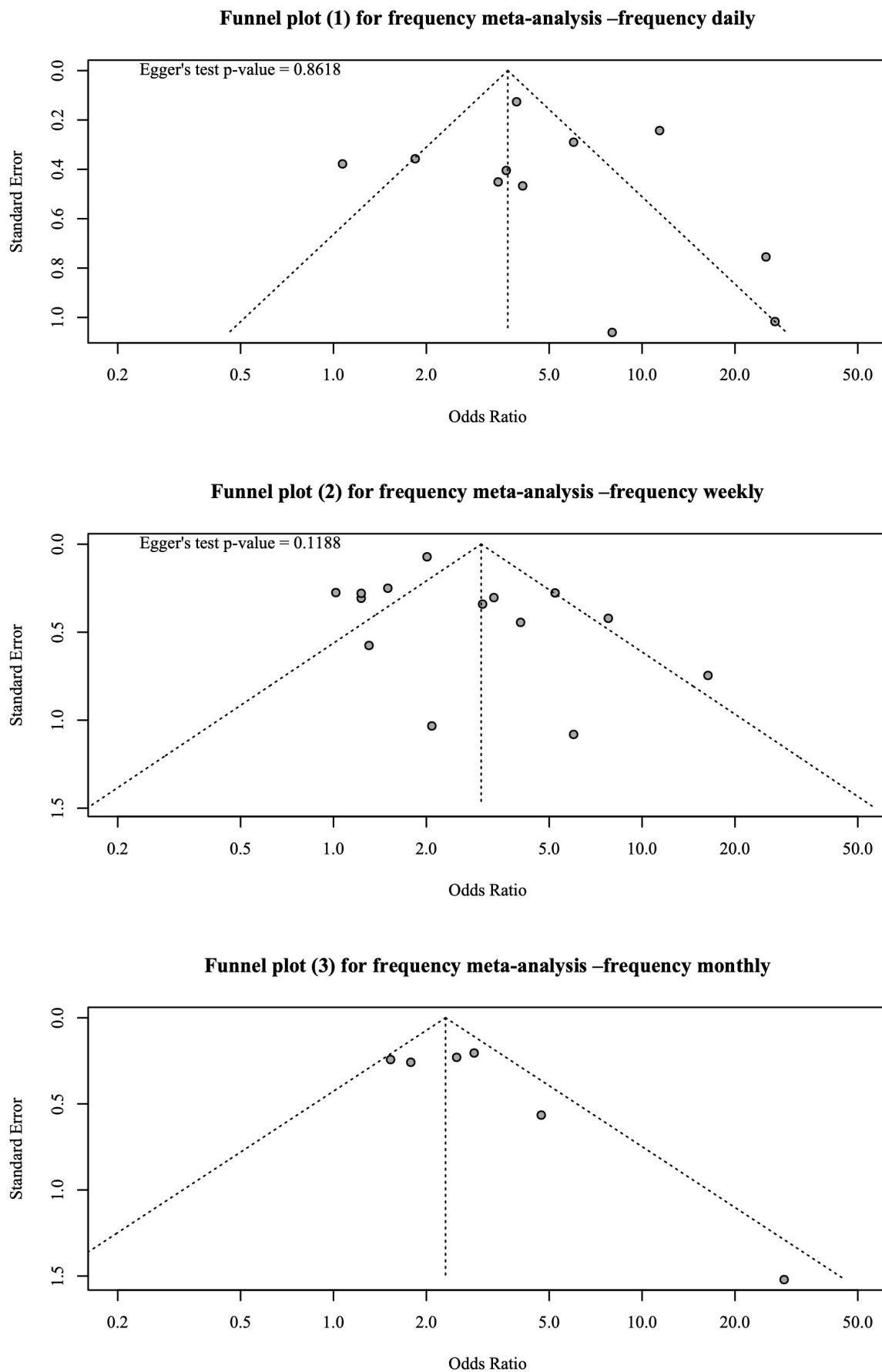


Figure S2.8. Egger's test p -value reported for all categories with $N \geq 10$.

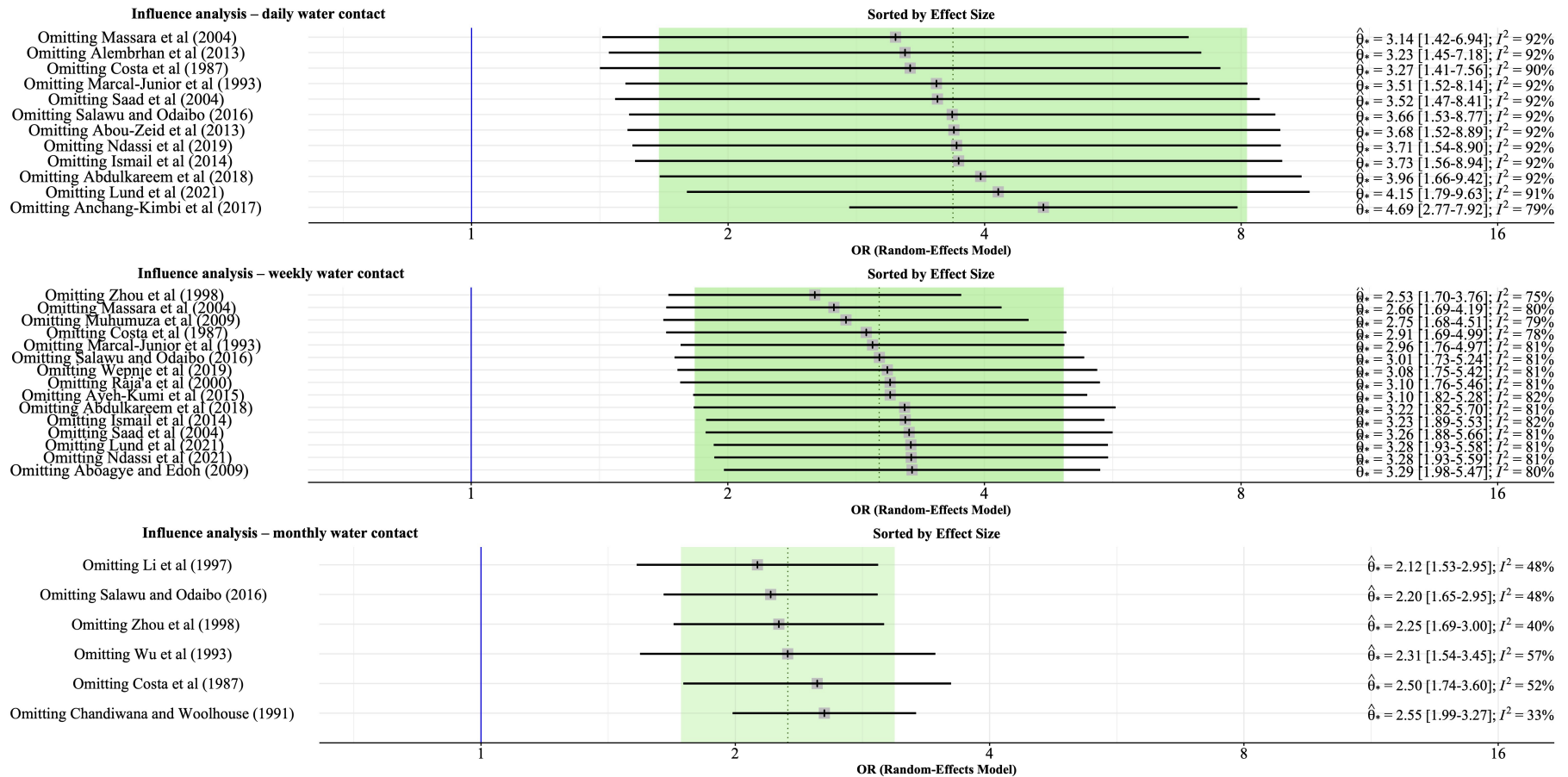


Figure S2.9. Influence analysis of any single study in the meta-analysis of the association between daily, weekly, monthly water contact and odds of schistosome infection.

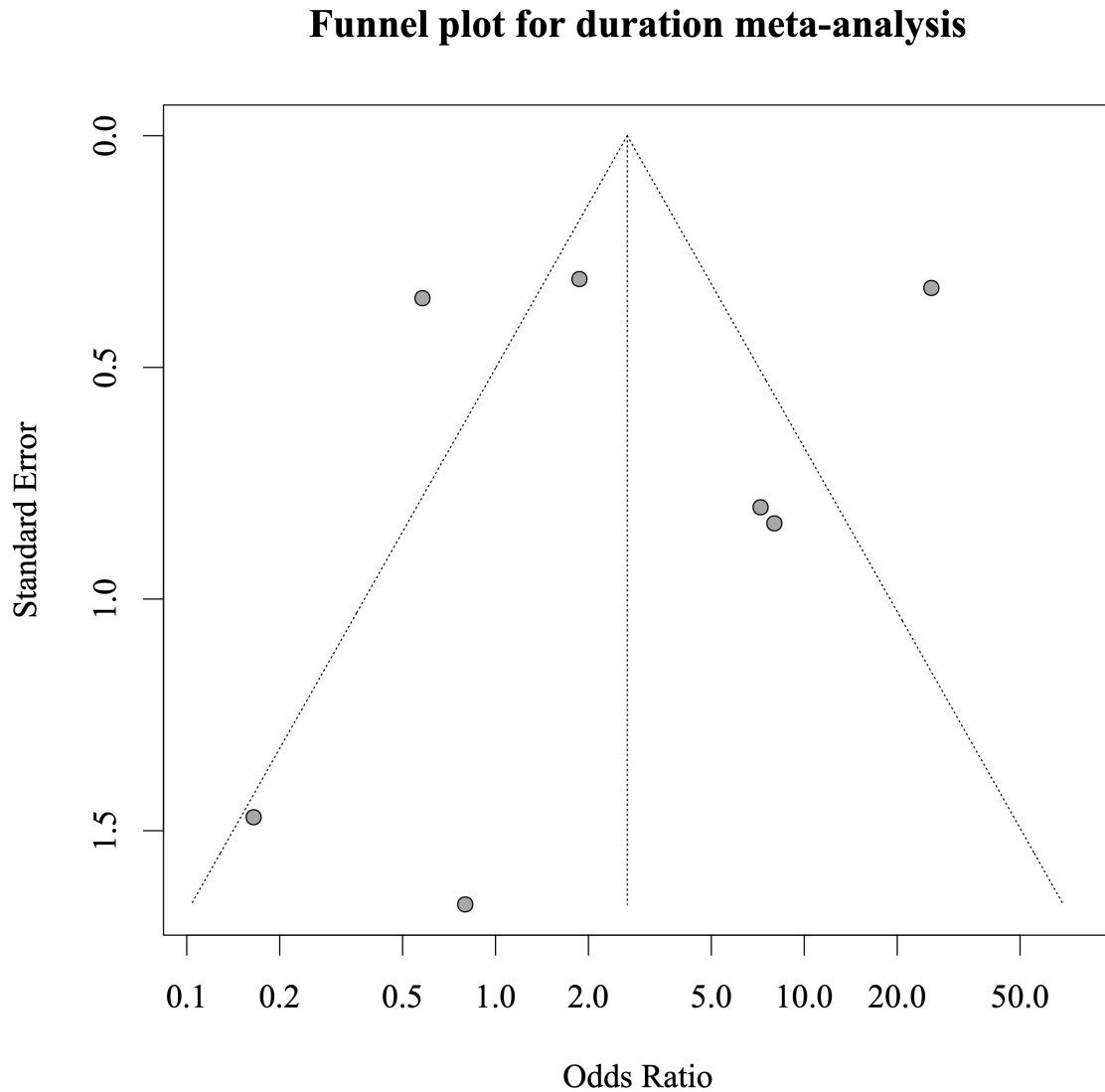


Figure S2.10. Funnel plot for meta-analysis of the association between individual-level water contact duration and odds of schistosome infection. No Egger's test was conducted due to the small number of studies ($N = 6$) because Egger's test may lack statistical power to detect bias when $N < 10$.

Funnel plot for activity meta-analysis

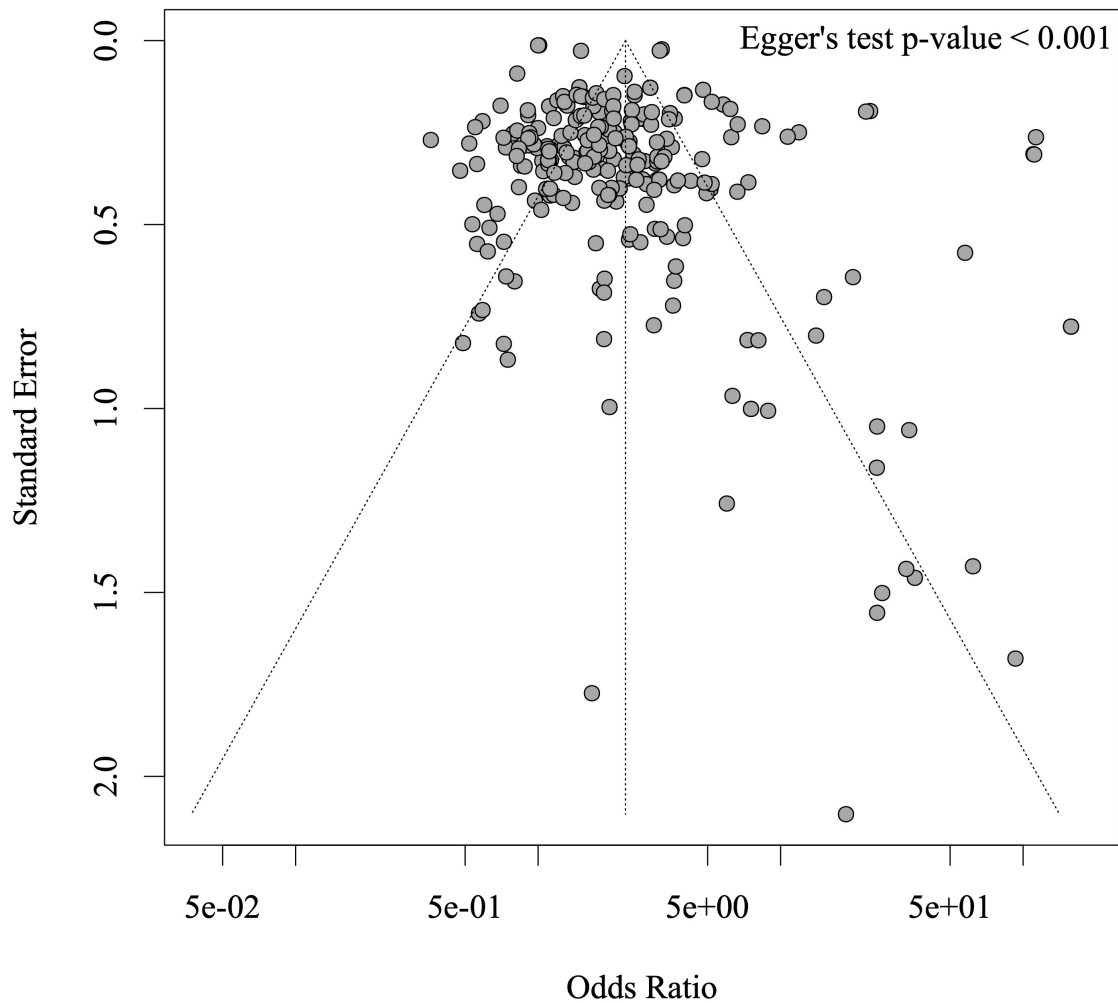


Figure S2.11. Funnel plot for meta-analysis of the associations between water contact activities and odds of schistosome infection. The Egger's test indicates significant funnel plot asymmetry ($p < 0.001$). The funnel plot includes effect sizes and standard errors from all water contact activities reported across studies, using a multilevel model clustered by study.

CHAPTER 2. SYSTEMATIC REVIEW AND META-ANALYSIS

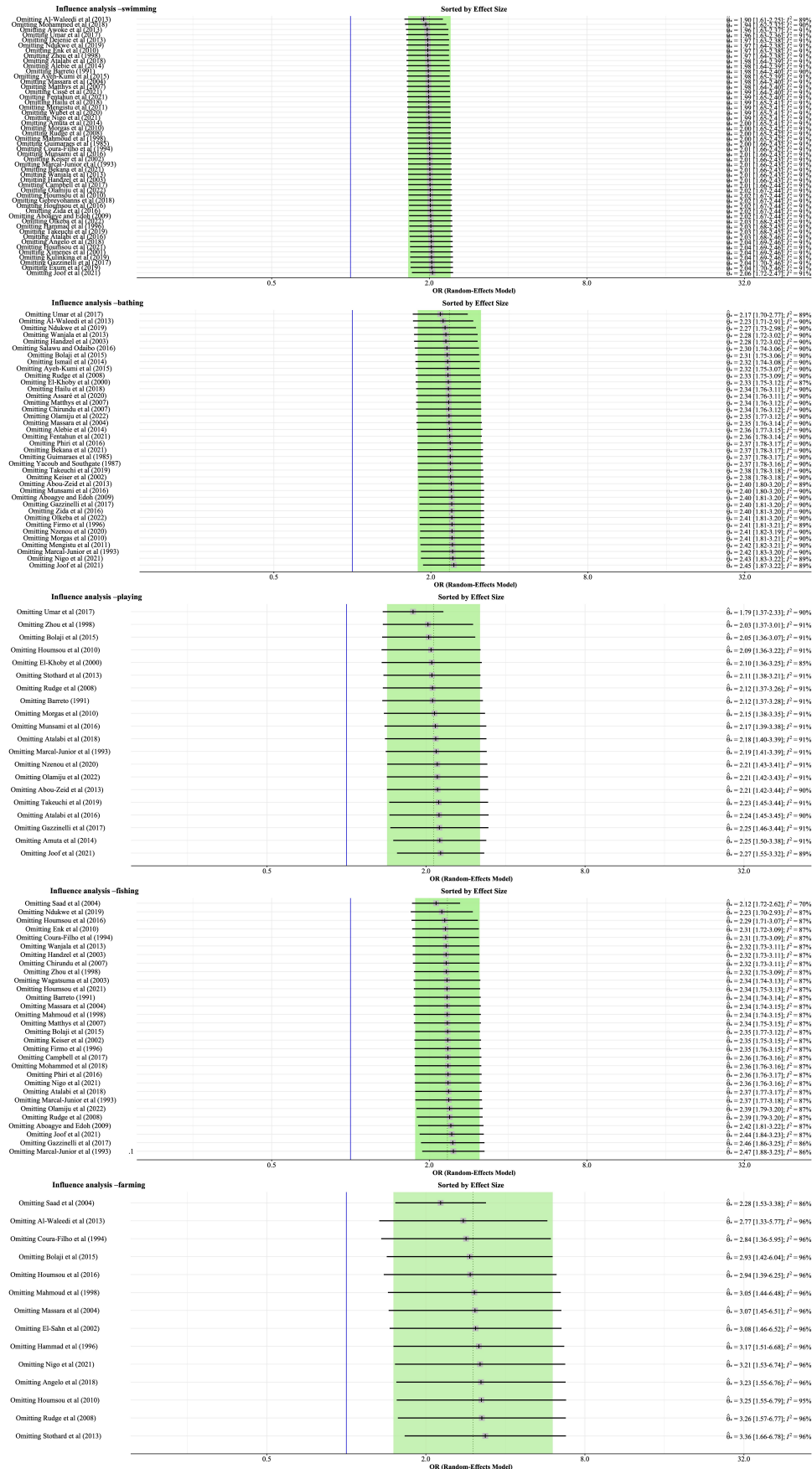


Figure S2.12. Influence analysis of any single study in the meta-analysis of the associations between water contact activities and odds of schistosome infection ($I/2$).

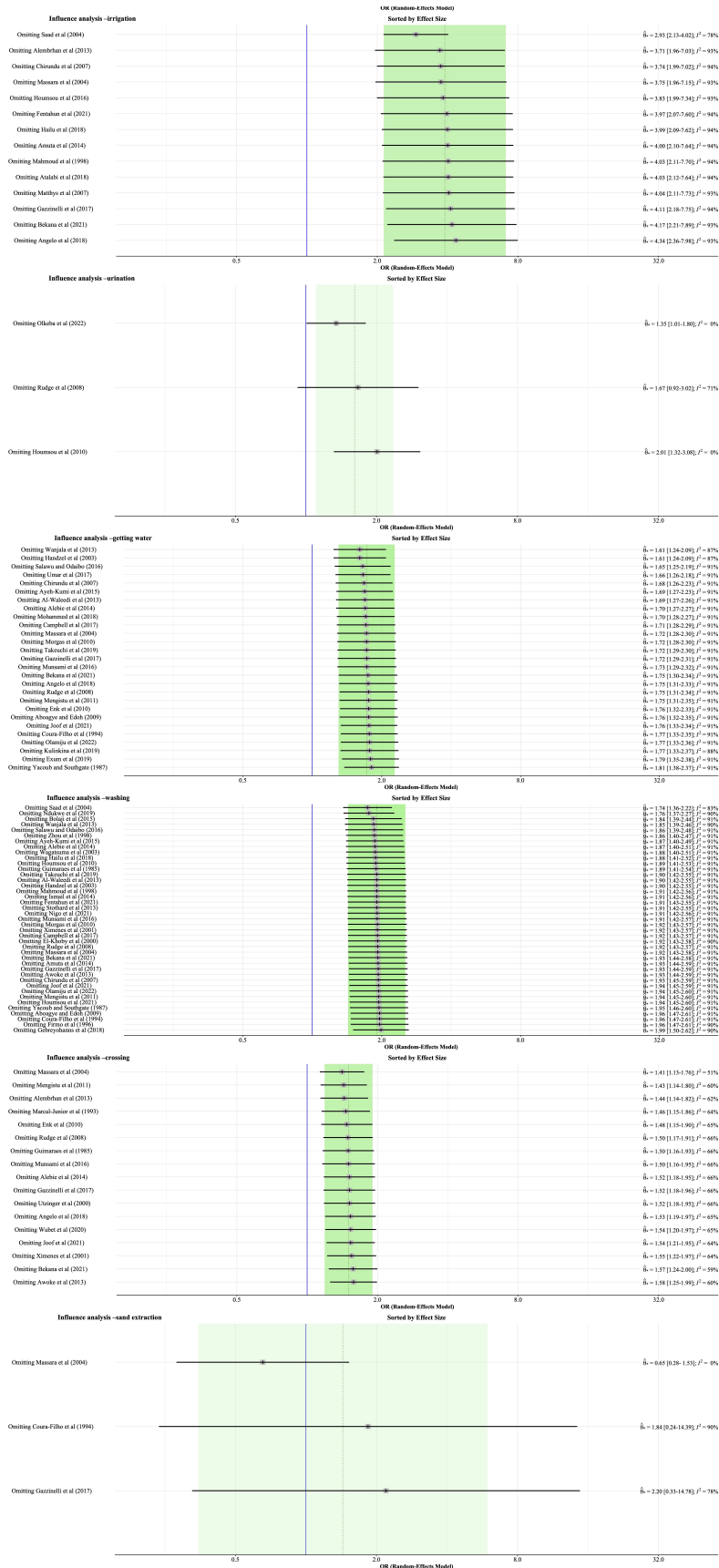


Figure S2.13. Influence analysis of any single study in the meta-analysis of the associations between water contact activities and odds of schistosome infection (2/2).

Chapter 3

A Data-Driven Analysis of the Determinants of Current Water Contact and *S. mansoni* Infection in the SchistoTrack Study

This chapter has been published as:

1. **Reitzug F**, Kabatereine NB, Byaruhanga AM, Besigye F, Nabatte B, and Chami GF. Current water contact and *Schistosoma mansoni* infection have distinct determinants: a data-driven population-based study in rural Uganda. *Nature Communications* 2024 Nov; 15. doi: 10.1038/s41467-024-53519-4.

Author contributions: Conceptualisation: FR and GFC. Data curation: FR, NBK, AMB, FB, BN, and GFC. Formal analysis: FR. Funding acquisition: GFC. Investigation: FR and GFC. Methodology: FR and GFC. Project administration: NBK and GFC. Resources: NBK and GFC. Software: GFC. Supervision: GFC. Validation: FR. Visualisation: FR. Writing – original draft: FR. Writing – review & editing: FR, NBK, AMB, FB, BN, and GFC.

3.1 Background

This chapter draws upon SchistoTrack cohort data from rural Uganda to investigate the complex drivers of schistosome transmission which include human behaviour, accessibility of safe water and sanitation, occupation, and ecological conditions for snail intermediate hosts. As exposure to schistosomes occurs during water contact with infested waterbodies such as lakes, rivers, or streams through various water contact activities, there is a need to capture both the human and environmental factors which determine exposure. Water contact has become a well-established proxy indicator for exposure due to the difficulties in directly measuring cercarial

exposure. However, among the 101 included studies in the systematic review in Chapter 2, only 21.8% (22/101) collected snail abundance data to account for environmental factors relevant for translating water contact into parasite acquisition risk.

Despite the need to integrate water contact with environmental variables, studies which created exposure indices have not necessarily found that this increased the ability to predict infection outcomes [1–3]. This may be due to limitations in data such as rough estimates of bodily immersion [1, 3], assumptions made about cercarial density [1, 2], or methodological limitations including no systematic variable selection and no out-of-sample validation [4]. Due to regular MDA, studying the role of exposure for infection is further complicated by immunity which can be acquired through past infection or successful treatment with praziquantel [5–7]. Understanding exposure also has proven difficult as no standardised exposure measurement tools exist and there is high heterogeneity across existing studies [4].

This chapter addresses the lack of comparative studies on predictors of exposure versus current infection. It has been found that infection prevalence varies based on household distance to waterbodies [8, 9], but whether this is explained by corresponding trends in water contact over distance remains unclear. Applied statistical models of exposure have predominantly focused on water contact as a predictor of infection without characterising water contact as its own outcome [4]. For estimating the force of infection, current mathematical models infer age-specific water contact levels from infection trends over age [10, 11]. These models typically assume that age-specific infection levels are a non-linear function of cumulative water contact up to that age and the survival of adult schistosomes within the host. As highlighted in Chapter 2, there is a need to characterise key determinants of water contact and assess whether infection and water contact follow similar trends over age.

The aim of this chapter is to comprehensively characterise water contact patterns in the SchistoTrack cohort using applied statistical models. The main research questions are: What are the major human-environmental determinants of water contact, and at which level (individual, household, or village) is the influence of those determinants? What are (if any) the shared dimensions of water contact determinants with infection determinants?

3.2 Methods

3.2.1 Study Setting, Participant Sampling, and Assessments

This chapter draws upon data collected as part of the SchistoTrack study, led by GFC at the University of Oxford and NKB, based at the Vector Control Division (VCD) of the Uganda Ministry of Health. I was part of the study team conducting the baseline data for the SchistoTrack cohort [12]. These data were collected in January and February 2022 during the dry season when human water contact and snail abundance were high [13]. Thirty-eight diverse rural villages were sampled across three districts of Pakwach, Buliisa, and Mayuge. We sampled villages within 3 kilometres (km) distance based on the village centre location reported by the local chairman of the shorelines of River Nile, Lake Albert and Lake Victoria (Fig. 3.2b–d). In all study districts, MDA with praziquantel has been carried out since the year 2003, targeting all children aged ≥ 5 years. Thirteen or more rounds of treatment have been administered to-date in each district. The most recent MDA was conducted over one year before this study. We randomly sampled a total of 1,459 households, approximately 40 per village, from village registers or MDA records. Sample size was determined as reported in [14]. All households with at least one child and one adult residing in the village for at least six months of the year were eligible. Informed consent was obtained from all participants. Parents or guardians consented on behalf of children < 18 years of age. Adults provided written consent for children and where possible children also provided written consent. All children provided informed verbal assent. After obtaining consent, questionnaires were administered to obtain socio-demographics, biomedical variables, WASH and environmental variables, and water contact patterns. At the end of the household interview, one adult (aged ≥ 18) and one child (aged 5–17) per household were selected for clinical assessments by the household head. All participants were treated for schistosomiasis using praziquantel following clinical examinations irrespective of schistosome infection status.

We also conducted village-level infrastructure and malacology surveys. For village-level infrastructure, information on the number of functioning public latrines and public taps/boreholes and their respective GPS locations as well as the GPS locations of all primary schools were

collected. We refrained from calculating community-level WASH coverage indicators as this could have risked introducing ecological fallacy into associations between WASH and infection (see Ref. [15] and reviewer comments in Ref. [16]). All latrines in the study area were pit latrines, constructed close to the shoreline, particularly in Buliisa and Pakwach, where some were destroyed during flooding in 2021—2022. Taps/boreholes were typically located slightly further from the shoreline than latrines and were free from contamination. To assess environmental risk, trained malacologists from the VCD mapped all water sites in the study area, guided by the village chairman or a village health worker, and recorded the GPS location of each site, site type, and the observed presence of human faeces at all water sites within the village. Malacologists collected all living and recently dead snails at each water site for 30 minutes, using both scooping and searching for snails on floating vegetation. Snail infectivity was established by exposing snails to natural sunlight for a maximum of eight hours the day after collection and determining the number of snails shedding human schistosome cercariae by water site.

3.2.2 Fieldwork During the DPhil

As part of the DPhil, I was extensively involved in the SchistoTrack baseline fieldwork. I supported preparation, data collection, and management on the ground, and helped with data processing and cleaning after I came back from Uganda. The extensive contributions of others to the fieldwork are acknowledged in the introduction in Chapter 1. Leading up to the fieldwork, I was in charge of preparing the GPS logger study (described in detail in Chapter 4). I supported implementing changes to the questionnaires already programmed in Open Data Kit (ODK) by OL and wrote quality control scripts for the clinical data. I also wrote code to generate participant barcode lists, which would be used by field staff to ensure selected participants attended all the different clinical stations. I was responsible for creating unique identifiers for the approximately 3,000 study participants. This involved devising a suitable barcode system, designing the barcodes in dedicated open-source software, and printing and laminating all 3000 barcodes in Oxford prior to the fieldwork. I also created spare barcodes and ensured we had a laminator and printer in the field to reprint barcodes, as needed.

During my two months in the field in January and February 2022, I helped with data collection and data management. Together with ME and initially GFC, I collected village infrastructure data; that is, I physically visited and recorded the GPS locations of all public taps/boreholes and latrines in each of the 38 study villages, identified the village centre guided by the chairman, and established where the village boundaries were. Additionally, I visited all primary schools which people in the study villages attended and recorded information on the number of pupils, catchment villages, grades offered, and MDA rounds and medicines available at the school level. I also visited all local health centres to establish their catchment populations and the available medicines. I visited all local drug shops (both formal and informal) and recorded the ten most used medicines and their respective prices. In terms of the clinical data collection, I generated barcode lists to ensure recruited patients attended all stations and ran quality control scripts in the field to help the field teams collect data from participants who had missed stations. I was responsible for the clinical data management of the Lumify point-of-care ultrasound application, I created daily backups of the data on field laptops and later transported all micro-USB cards back to Oxford. For direct water site observation, together with ME and initially GFC, I trained locally recruited direct observers, handed out the observation forms, and later helped to collect filled-in forms and assisted with digitising the data.

Post-fieldwork, I helped with data cleaning, especially for the clinical data and village infrastructure data, creating the final datasets made available to the Chami Group cluster on the BMRC cloud. I also created village notes which consolidated the handwritten observations that ME, GFC, and I had made during the fieldwork. I was also involved in creating documentation for all datasets to provide details on how the data was collected, any known data issues, and instructions for future users.

In January 2023, I went to Uganda for one week to help digitise some additional information from the water contact observation forms and assist with training the Ugandan team at the VCD with a training session on Geospatial Information Systems (GIS), led by JL.

3.2.3 Water Contact Outcome

To ascertain water contact information in SchistoTrack, survey questions were administered to the household head and/or wife to elicit information on 11 different water contact activities conducted by household members, and their typical weekly frequency, duration, and time of day. The 11 activities were: getting drinking water, washing clothes with/without soap, bathing with/without soap, washing jerry cans or household items, washing clothes, collecting papyrus, fishing, fish mongering, collecting shells, and swimming or playing. (For self-reported soap use, any type of soap was considered eligible. Both bar soap and powder soap were used locally.) The choice of activities was informed by the literature and knowledge of the VCD team to ensure all common water contact activities in the study area were captured. The typical duration per trip to waterbodies for each activity in the household was elicited using predefined duration categories (< 30 minutes, 1 hour, 2 hours, 3 hours, $\geq$ 4 hours). For frequency, the typical number of trips per week in the household was collected. The typical time of day for each activity was elicited using the following categories: before sunrise (1–6 a.m.), after sunrise (6–9 a.m.), late morning (9 a.m.–12 p.m.), midday (12–3 p.m.), late afternoon (3–5 p.m.), early evening (5–7 p.m.), after sunset (7–9 p.m.), late evening (9 p.m.–1 a.m.).

The main outcome was having any water contact which was coded as binary and defined as an individual engaging at least in one water contact activity per week. We chose having any water contact as the primary representation of exposure based on Chapter 2 where it was shown that having any water contact was associated with infection status. Secondary outcomes included the type of water contact activity and the frequency and duration of water contact. For activity types, we grouped water contact into three distinct categories: domestic, occupational, and recreational water contact. Domestic water contact was defined as engaging in getting drinking water, washing clothes with/without soap, bathing with/without soap, washing jerry cans or household items, or washing clothes. Occupational water contact was defined as engaging in collecting papyrus, fishing, fish mongering or collecting shells. Recreational water contact was defined as swimming or playing. To obtain individual-level water contact frequency, we summed the self-reported weekly number of water contacts across all activities for each individual. Duration was obtained by multiplying the weekly frequency of each activity with

its duration and summing durations across all activities for each individual.

We also generated an additional 82 granular water contact variables indicating the frequency and duration for each of the 11 water contact activities, the time of day of each activity, and whether an activity was conducted at a high-risk time (10 a.m. – 3 p.m.), were included. We also included variables for the frequency and duration of domestic, recreational, and domestic activities and variables indicating whether anyone in the household conducted any of the 11 water contact activities. These more detailed variables were only used in (Fig. 3.9b)

3.2.4 *S. mansoni* Infection Outcome

Kato–Katz stool microscopy and point-of-care circulating cathodic antigen (POC-CCA) diagnoses were completed for schistosomiasis as described in [14]. We prepared duplicate, thick smear slides from a single, stool sample produced on the morning of the slide preparations. Senior technicians re-read a random sample of 10% of all slides for quality control. Readings were converted to eggs per gram (EPG) of stool. Infection was defined as > 0 EPG. In accordance with the current WHO schistosomiasis control guideline [17], heavy infection intensity was defined as ≥ 400 EPG. We chose infection status (based on Kato–Katz diagnosis) as the primary infection measure as infection status is relevant for transmission and prevalence has been used to monitor reductions in transmission [18] and identify persistent hotspots [19, 20]. POC-CCA was not chosen as the primary infection measure due to its lack of specificity [21] and undemonstrated relevance for transmission [22]. POC-CCA was only used in sensitivity analyses to demonstrate the robustness of main models in the potential scenario of missed light infection intensities. To this end, individuals who were Kato-Katz negative but POC-CCA positive (positive 1–3 band) were recoded as positive. As POC-CCA detects worm antigens, this recoding based on POC-CCA may help to identify individuals with nascent infection, single sex infections or post-menopausal infections not picked up by Kato-Katz [23].

3.2.5 Individual, Household, and Village Human–Environmental Variables

Detailed definitions of all variables are provided in Table S3.1. We curated a set of 33 candidate variables for our main analysis and aimed at comprehensively covering key human-

environmental determinants of water contact and infection (27 variables for the water contact models and six additional water contact variables for the infection models). At the individual level, we included self-reported age, gender, occupation, and highest education level completed as covariates, based on their relevance for water contact and infection [24]. Age was coded to the nearest year and was a continuous variable as grouping age into categories could lead to information loss [25]. We generated an age^2 term to model non-linear associations of water contact and infection.

Household-level variables were as follows. To account for the influence of socio-economic status, we generated a home quality score variable as described in [26]. WASH access may influence water contact by crowding out open water contact or may affect infection via reducing faecal contamination and thus environmental risk [27]. We generated sanitation variables indicating the Euclidean distance from the household to the closest public latrine and presence of a household home latrine. Having a flush or pour-flush toilet, a covered pit latrine with/without privacy, or a composting toilet was considered as having a home latrine. Households who reported open defecation or having only bucket latrines were considered as not having a home latrine, as per WHO–UNICEF Joint Monitoring Committee 2017 methodology guideline [28]. We also generated water infrastructure variables indicating the Euclidean distance to the closest tap/borehole, household water purification method used (if any), the presence of a handwashing facility, and whether the household used public taps/boreholes as their primary source of drinking water. Overall, 52.8% of households (95% CI 51.0–54.6%) used public taps/boreholes as their primary source of drinking water. We also generated variables for the number of latrines, and the number of taps/boreholes per village.

At the village level, we recorded a wide range of characteristics for water sites, of which some were combined to create distance metrics. The types of water sites (river, marsh, beach, swamp, pond) in each village were recorded by malacologists. The number of water sites per village was calculated from water sites mapped by malacologists. To measure faecal contamination, malacologists observed at all sampled water sites whether there was human stool present. Past research has shown the existence of lake gradients in infection [8, 9, 29]. We therefore generated Euclidean distance variables to assess whether these gradients were explained by variation in

water contact behaviour across lake distance. Variables were defined as follows: Household lake distance was the Euclidean distance of the household GPS location to the closest water site mapped by malacologists. Village lake distance was computed by assigning households to the village centre reported by the local chairman and calculating the Euclidean distance to the closest water site. For school lake distance, all children attending primary school were assigned to the school closest to the household. We then calculated the Euclidean distance from the school to the closest water site. Distance to the type of water site closest to the household was generated using information on the type of site collected by malacologists.

The information we collected about intermediate snail hosts and infection prevalence was also incorporated at the village level. The snail variables included were the number of water sites with infected snails per village, number of snails within 1 km of the household, number of snails per village, number of infected snails per village, presence of any infected snails in the village, distance to the closest water site with infected snails, and whether the closest infected water site from the household was within the village. Village-level prevalence was defined as per WHO schistosomiasis control guidelines with 10–49% prevalence by Kato–Katz corresponding to moderate and $\geq 50\%$ prevalence corresponding to high endemicity [17]. There were no low-prevalence ($\leq 10\%$) villages in our study.

3.2.6 Comparison of Self-Reported Water Contact with Direct Water Contact Observations

Direct water contact observations were conducted in a subset of 12 villages concurrently with the collection of self-reported water contact data. All water contacts at two water sites per village were recorded by trained local village health team members who were trusted by the communities and knew all individuals within their village. Contact was observed between 6 a.m. and 6 p.m. for two weeks and each visitation of the water site, defined as having at least one body part immersed, was recorded as a separate water contact event. Age group, as defined per recent WHO schistosomiasis control guideline [17], as well as gender, activity, bodily immersion, time, and duration were recorded. By comparing the age distribution in direct observation data for participants aged ≥ 5 years with the age distribution of water contact

in self-reported data from the same 12 villages, we assessed the extent to which survey data were representative of the age composition in the observation data.

3.2.7 Geospatial Data

Waterbody boundaries shown in Fig. 3.2 were from a public shapefile of waterbodies in Uganda, made available online under a Creative Commons Attribution (Non-commercial 4.0 International) license by the World Resources Institute [30].

3.2.8 Statistical Analysis

All statistical analyses were conducted in R version 4.1.0. To describe the univariate non-linear trends of water contact and infection over age, we used generalised additive models (GAMs) from the `mgcv` package [31]. The number of knots was separately chosen for each GAM based on the `mgcv` diagnostic function `k.check`. The GAM results informed our subsequent modelling of age in the regression models to predict water contact and infection where we used age and age² terms to account for non-linearities. We also modelled trends in water contact over distance to the nearest water site using GAMs and obtained distance gradients from these GAMs by computing the average slope over household distances between 0–1 km. We refrained from reporting distance gradients ≥ 1 km household distance, as only 4.7% of participants (137/2867) lived in households ≥ 1 km from a water site. We used Spearman correlation coefficients (ρ_s) to assess correlations between water contact and infection status across participants. We also assessed pairwise correlations of these variables within households. All *t*-tests conducted in this chapter were two-sided. To select predictors of water contact and infection status, respectively, the same variable selection procedure was applied separately using the same list of input covariates except for six additional water contact variables which were only included in the infection status models. We included a district-level fixed effect in every model to account for unobserved district-level differences in water contact and infection. To assess the relevance of snail variables for water contact and infection as well as the relevance of more granular water contact variables for predicting infection, we used Bayesian Variable Selection (BVS) to select additional predictors from the sets of seven snail variables and 82 granular water contact variables, respectively,

while constraining all previously selected variables to remain included. This way, we tested whether adding snail variables or more granular water contact variables improved predictions compared to the models with our main sets of predictor variables. BVS was implemented using the Bayesian Adaptive Sampling (BAS) package in R [32]. The outputs of this analysis were a ranking of the best-fitting models with associated posterior probabilities for each model being the best model as well as marginal inclusion probabilities bounded between 0 and 1 for each variable, indicating the importance of including each of the predictors in our water contact and infection models, respectively. The large number of candidate variables yielded a prohibitively large model space (2^{27} possible models for water contact). Therefore, we used Markov Chain Monte Carlo (MCMC) methods implemented within the BAS package to sample models based on their posterior probabilities and avoid enumerating all possible models [33]. We ran 2×10^7 MCMC iterations, a number deemed sufficient based on in-built MCMC diagnostics of the BAS package. The prior distribution of the models was set to a uniform distribution which assigned equal probabilities to all models. For the distributions on the regression coefficients, we used Jeffreys–Zellner–Siow priors which have been shown to have desirable mathematical properties and which provide more consistent results under both the null and the alternative model than alternative priors such as hyper-g priors or Empirical Bayes priors, and are also more consistent than non-Bayesian methods such as the Akaike Information Criterion [34–36]. To build the final water contact and infection models, we selected all variables with inclusion probabilities of $pr \geq 0.5$ in the BVS output which yielded the so-called median probability model [37]. The choice of $pr \geq 0.5$ was motivated by research showing that the median probability model, not the highest probability model, often yields the best predictive performance [33]. As an alternative to BVS, we used likelihood-ratio tests (LRTs) with a cut-off of $p < 0.05$. For LRTs, we compared a model with just one variable at a time against an empty model with village-level fixed effects only to select predictors of water contact and infection. A comparison of the variable sets selected using Bayesian variable selection and LRTs is shown in Table S3.7. For models shown in Fig. 3.9b–c, we included the additional snail variables and water contact variables, described above (model results shown in Figs. S3.15 and S3.16).

After variable selection, we used generalised linear models with a binomial family and a

log-link function to predict whether an individual had any water contact ($n = 2867$). For all regression models, we clustered standard errors at the household level using the `sandwich` package in R [38] to account for our sampling design where households were randomly selected, but not individuals. We considered the use of both random and fixed effects models to account for unobserved household and village-level confounders. Using a Durbin–Wu–Hausman test [39], we compared the consistency of village-level random effects models against an alternative model using district fixed effects. The test statistic from the Durbin–Wu–Hausman test comparing the water contact random effects model with the respective fixed effect model provided strong evidence ($\chi^2 = 137.46$, $p\text{-value} < 0.01$) that random effects models would have provided inconsistent estimates. We also ran the Durbin–Wu–Hausman test for the infection status model and found that a random effects model would have provided consistent estimates ($\chi^2 = 3.6994$, $p\text{-value} = 0.99$). Yet, when we analysed model residuals from the fixed-effects infection status model, we found that the intra-class clustering coefficient (ICC) was 0.017, indicating that there was very little unaccounted household-level variation. Therefore, we believe that the use of district-level fixed effects was justifiable both for modelling water contact and infection status.

To identify which variables predicted a specific type of water contact, we used the main set of selected variables from the water contact model to predict domestic, occupational, and recreational water contact, respectively, in separate logistic regression models. In these models, we removed all participants engaging in multiple types of water contact activities to generate mutually exclusive categories. This removal resulted in 6.3% individuals (183/2867) being excluded. To adequately model overdispersion of water contact frequency and duration, we used negative binomial models to predict frequency and duration using all selected predictors of water contact. For all negative binomial models, we tested whether they adequately modelled zero-inflation using the `check_zeroinflation` function from the `performance` package in R [40]. For infection, we used the selected infection variables from Bayesian variable selection in logistic regression models to predict infection status as well as heavy infection ($n = 2867$). We modelled frequency and duration of water contact as well as infection intensity using negative binomial regression models. Across all models, we computed generalised variance inflation factors (gVIFs) to diagnose multicollinearity [41] but did not remove any variables as none

had $gVIFs > \sqrt{10}$. Ten-fold cross-validation was used to assess the predictive capacity of the statistical models with binary outcomes based on the area under the receiver operating curve.

3.3 Results

3.3.1 Water Contact Type, Timing, Frequency, and Duration

A study overview is presented in Fig. 3.1. Detailed variable definitions and characteristics of the 2867 participants are presented in Table S3.1. Among study participants, 46.7% (1339/2867, 95% confidence interval (CI): 44.9–48.5%) reported that they had water contact with open freshwater bodies at least once per week. The median frequency of water contact was six times per week (interquartile range (IQR): 3–11) and the median duration in hours per week was eight (IQR: 3.5–17.5). Both frequency and duration exhibited overdispersion, with only few individuals engaging in high-duration or high-frequency water contact (Fig. S3.1). The three

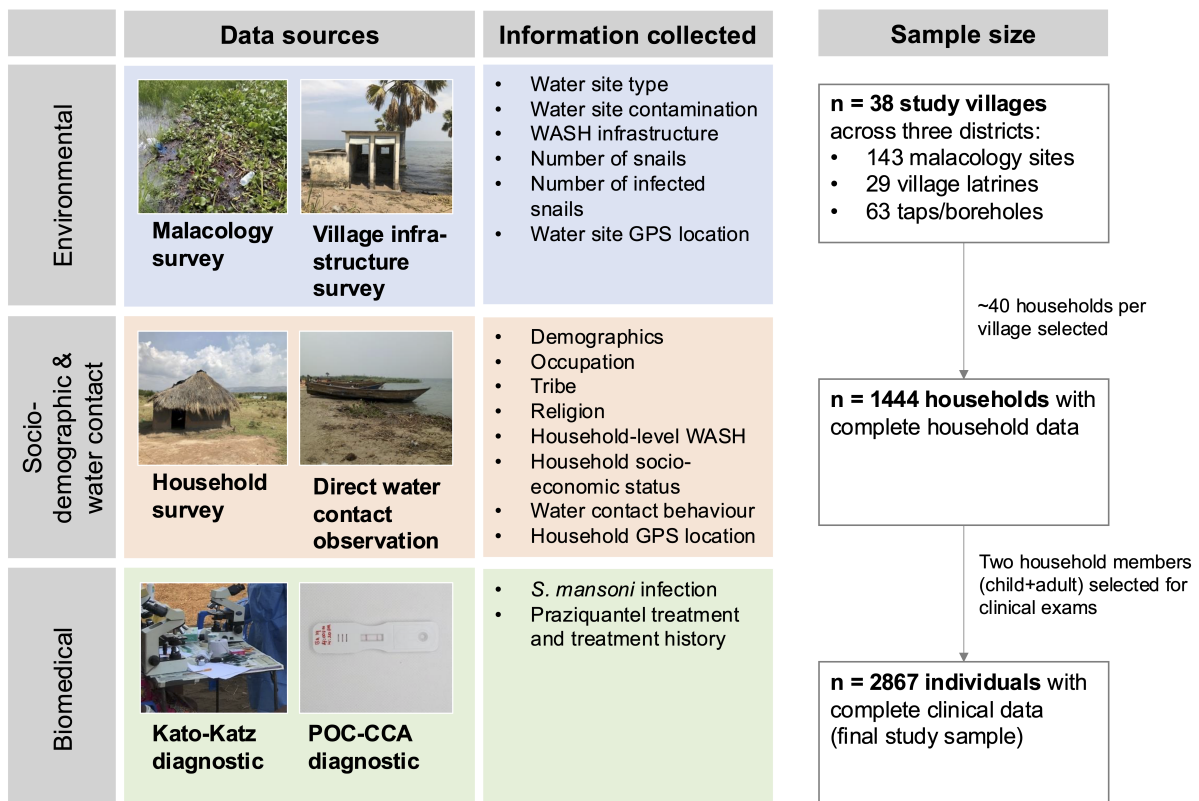


Figure 3.1. SchistoTrack study overview. Data sources, information collected, and sample size of the SchistoTrack study. Abbreviations: WASH = water, sanitation, and hygiene. GPS = global positioning system.

most common water contact activities among participants were getting drinking water (17.3%, 497/2,867), washing clothes with soap (16.8%, 481/2867), and fishing (12.3%, 354/2867, Table S3.3). People washing clothes with soap were as likely as other participants to report washing clothes without soap (5.8%, 28/481, versus 5.1%, 124/2386, $p = 0.67$). Overall, water contact was most commonly conducted between 6–9 a.m. with 23.3% (956/4046) of contacts taking place at this time. The typical time of day of water contact varied by activity type. Washing clothes with soap and getting drinking water were most commonly done after sunrise (6–9 a.m.) with 32.8% (163/497) and 34.1% (164/481) of water contacts related to these activities, respectively, occurring during this time. For fishing, the early evening (5–7 p.m.) was the most common time, accounting for 24.2% (86/354) of fishing water contacts. Concerning frequency of water contact, the median number of weekly trips was two for washing clothes with soap (IQR: 2–5), seven for getting drinking water (IQR: 5–14), and four for fishing (IQR: 2–7). Among these activities, the median duration of water contact was highest for fishing at four hours (IQR: 3–4). Washing clothes with soap or getting drinking water had a median duration of only one hour, noting 30 minutes as the minimum amount of time recorded in the study survey.

3.3.2 Water Contact Dependence on Distance to Water Sites

Households were located within 0.01–2.17 km (mean 0.35 km, 95% CI: 0.43–0.36 km) of the River Nile, Lake Albert or Lake Victoria (Fig. 3.2). Household distance varied by district. In Pakwach District, the mean distance to the River Nile or Lake Albert was 0.41 km (95% CI: 0.39–0.43 km); whereas in Buliisa District, the mean distance to Lake Albert was 0.24 km (95% CI: 0.23–0.26 km). Household waterbody distances were more similar between Mayuge and Pakwach, albeit they were the districts furthest apart. Water contact varied by waterbody distance; in households residing directly on the shoreline (≤ 100 m), 54% of individuals (315/589) had water contact. In households located >1 km from a water site, 26% (36/137) of individuals reported having water contact. Water contact declined rapidly with shoreline distance, as 80% of all individuals with water contact lived within 0.43 km Euclidean distance of water sites (Fig. S3.2). Among individuals within 0.43 km, 58.1% had water contact, whereas only 26.1%

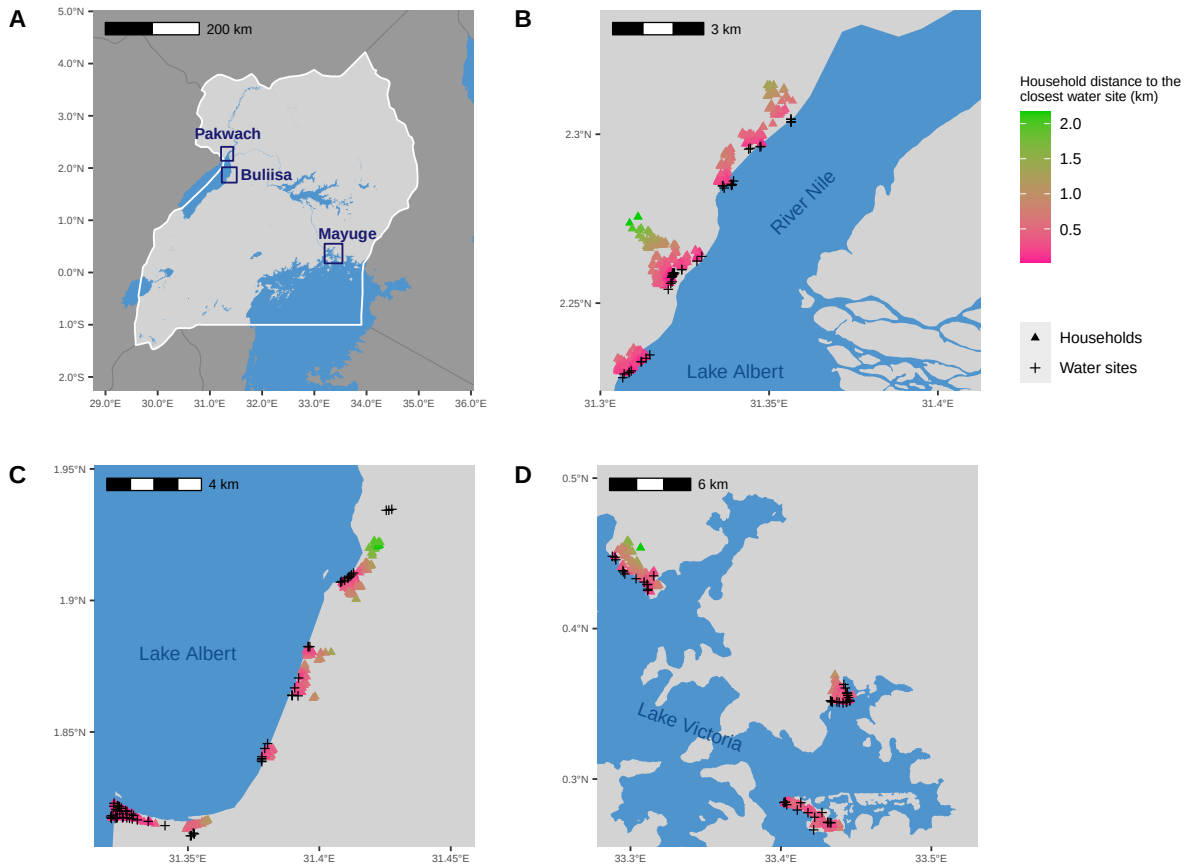


Figure 3.2. SchistoTrack study areas. The locations of 2867 study households (randomly displaced by 50m) and 143 water sites sampled by malacologists are shown. Waterbody boundaries were taken from the open resource dataset from the World Resources Institute. (A) Locations of the three study districts Pakwach, Buliisa, and Mayuge in Uganda. (B) Zoomed in view of Pakwach. Households and water sites are shown along Lake Albert and the River Nile. (C) Zoomed in view of Buliisa. Households and water sites are shown along Lake Albert. (D) Zoomed in view of Mayuge. Households and water sites are shown along Lake Victoria. Household distance in this figure represents Euclidean distance of the household location to the closest water site.

of individuals living further than 0.43 km away had water contact. Using GAMs, we modelled the decline in water contact over household distance to the closest water site. For every 100 m increase in distance between 0–1 km, a 3.4% point reduction in the proportion of participants with water contact was found. Consistently decreasing gradients were observed across domestic, occupational, and recreational activities (Fig. 3.3a). Gradients in water contact differed depending on the distance metric used (Fig. 3.3b). Village centre distance to the closest water site produced weaker gradients than household distance, with a 1.9% versus 3.4% point reduction for every 100 m increase with village distance compared to household distance (Fig. 3.3c).

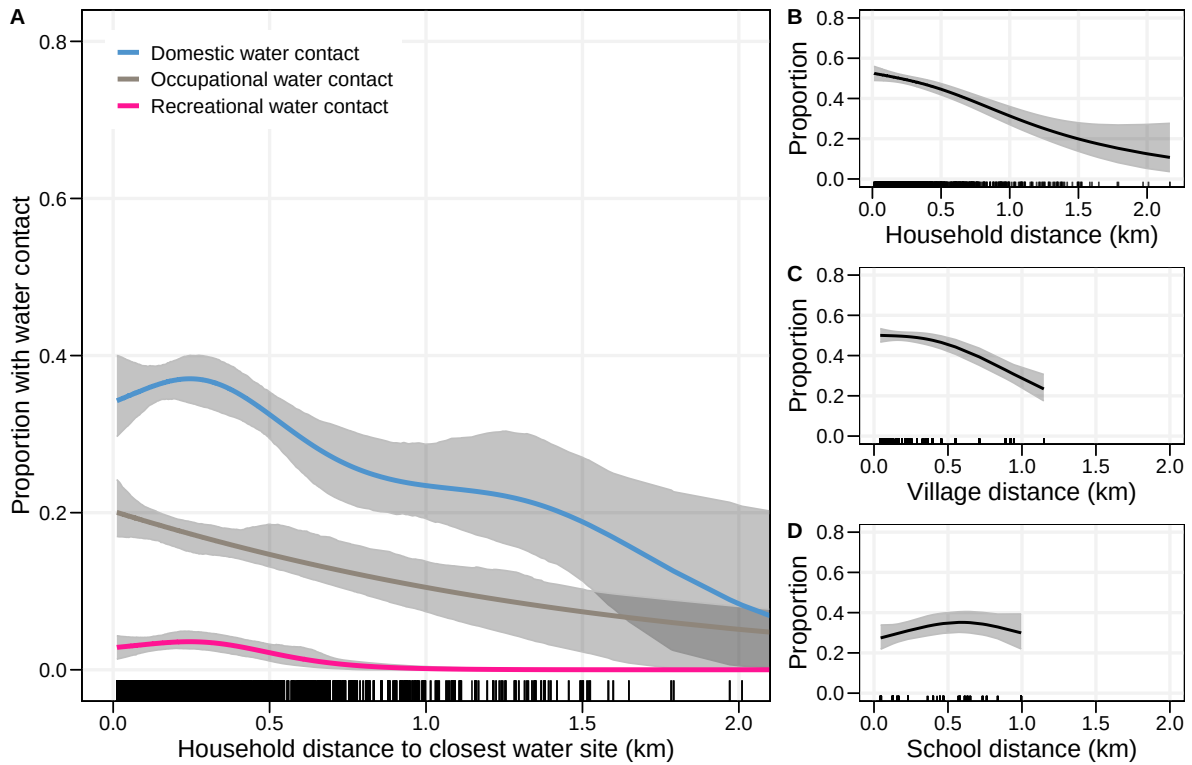


Figure 3.3. Proportion of individuals with water contact over household waterbody distance. The proportion of individuals with current water contact over Euclidean household distance to the closest water site (in km) was modelled using generalised additive models. (A) Proportion of individuals with domestic, occupational, and recreational water contact. (B–D) Comparison of lake gradients in water contact across alternative waterbody-distance metrics: household distance, village centre distance, and school distance to the closest water site. (B) Proportion of individuals with any water contact by household-level distance to the closest water site. (C) Proportion of individuals with water contact by village centre distance to the closest water site. Village centre location was identified as the cultural centre by the village chairman. All households within a village were assigned the distance from the village centre to the closest water site. Compared to household water site distance with a maximum of 2.2 km, the village centre distance distribution was truncated at 1.2 km as the furthest village centre in this study was 1.2 km away from the closest water site. (D) Prevalence of water contact by distance from the school closest to the household to the nearest water site. Sample restricted to $n = 597$ enrolled children. The 95% confidence intervals in Panel A were computed via bootstrap with 1000 repeats.

When using primary school distance to the closest freshwater body, there was no relationship between distance and the percentage of households who engaged in water contact (Fig. 3.3d).

3.3.3 Gender and Age-Specific Water Contact Patterns

Gender and age-specific patterns of water contact were examined using GAMs. Overall, 49.3% (775/1573) of females and 43.5% (564/1294) of males reported water contact. The frequency

of water contact was higher for females than males at a median of seven versus five trips per week (Wilcoxon rank sum test, $p < 0.01$, Table S3.4). Conversely, females had a lower duration of water contact than males, with seven compared to ten hours reported per week (Wilcoxon rank sum test, $p < 0.01$, Table S3.5). Relationships between current water contact with gender and age were non-linear and varied substantially over the life course. The modelled age-dependent proportion of participants with water contact was 18% (95% CI: 14–18%) at age five, peaked at 70% (95% CI: 66–74%) at age 30, and declined to 28% (95% CI: 20–36%) at age 70 (see Fig. 3.4a). Across ages 18–35, the proportion of individuals with water contact was higher for males than females (Fig. S3.3). Beyond any water contact, the type of water contact activity was gender-dependent (Table S3.3). For instance, 24.3% (315/1294) of males reported going fishing, while the same figure was only 2.5% (39/1573) among females ($\chi^2 = 311.56$, $p < 0.01$). Collecting drinking water was reported by 22.1% (348/1573) of females but only 11.5% (149/1294) of males ($\chi^2 = 55.02$, $p < 0.01$). Washing clothes with soap was also more

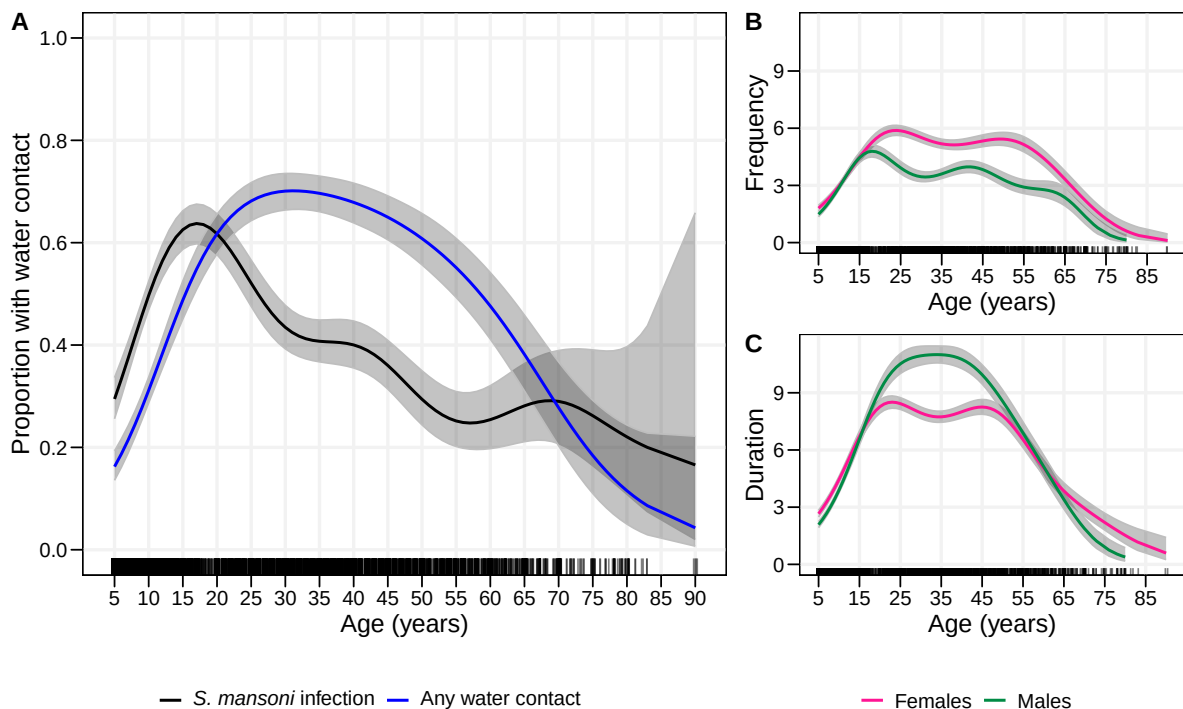


Figure 3.4. Variation in water contact and infection over age. (A) Proportion of individuals with current water contact and *S. mansoni* infection prevalence over age modelled using generalised additive models. (B) Frequency of water contact measured by the weekly number of trips to waterbodies over age and gender. (C) Duration of water contact measured in numbers of hours per week, over age and gender.

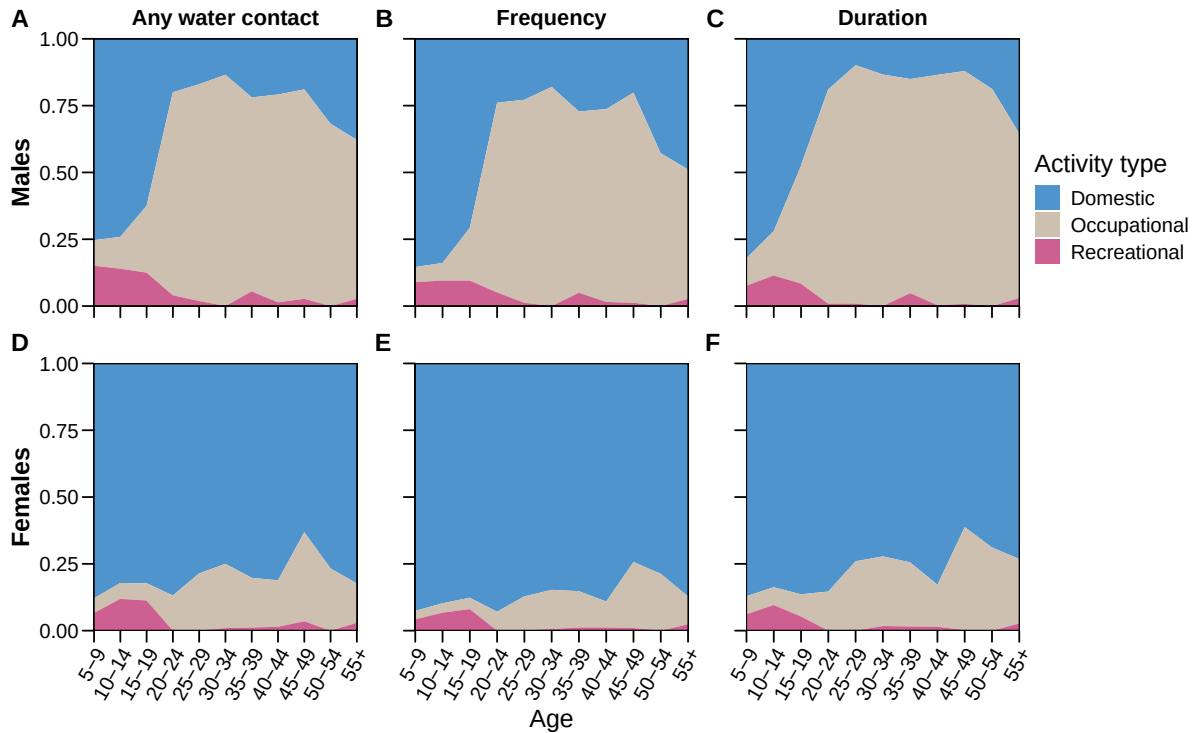


Figure 3.5. Composition of water contact activities over age and gender. (A) Proportion of all water contacts within each five-year age group which were domestic, occupational, or recreational over age for males. (B) Frequency, measured as the proportion of trips to waterbodies that were domestic, occupational, or domestic over age for males. (C) Duration, measured as the proportion of water contact time which was accounted for by domestic, occupational, or recreational activities over age for males. (D) Proportion of all water contacts within each five-year age group which were domestic, occupational, or recreational over age for females. (E) Frequency, measured as the proportion of trips to waterbodies that were domestic, occupational, or domestic over age for females. (F) Duration, measured as the proportion of water contact time which was accounted for by domestic, occupational, or recreational activities over age for females.

common among females than males (23.4% (368/1573) versus 8.7% (113/1294), respectively, $\chi^2 = 108.27$, $p < 0.01$). The proportion of water contacts taking place during peak cercarial shedding time, defined as between 10 a.m.–3 p.m. [42], was not significantly different between males and females for any activity. However, predominantly female activities—getting drinking water, washing clothes with soap, and washing jerry cans or household items—were more likely to be conducted during peak shedding time than male activities such as fishing (Table S3.6). Gender differences in water contact frequency, duration, and dominant activity type became pronounced in adolescence between ages 15–19 (Figs. 3.4 and 3.5). Until age 15, water contact patterns of both females and males included mostly domestic activities (Fig. 3.5). After age 15, females maintained high levels of domestic water contact, while the relative involvement of

males in domestic activities strongly declined (Fig. 3.5). Among adults (age ≥ 18), domestic water contact accounted for 75.3% of total water contact duration for females, while occupational water contact accounted for 81.8% of total water contact duration for males.

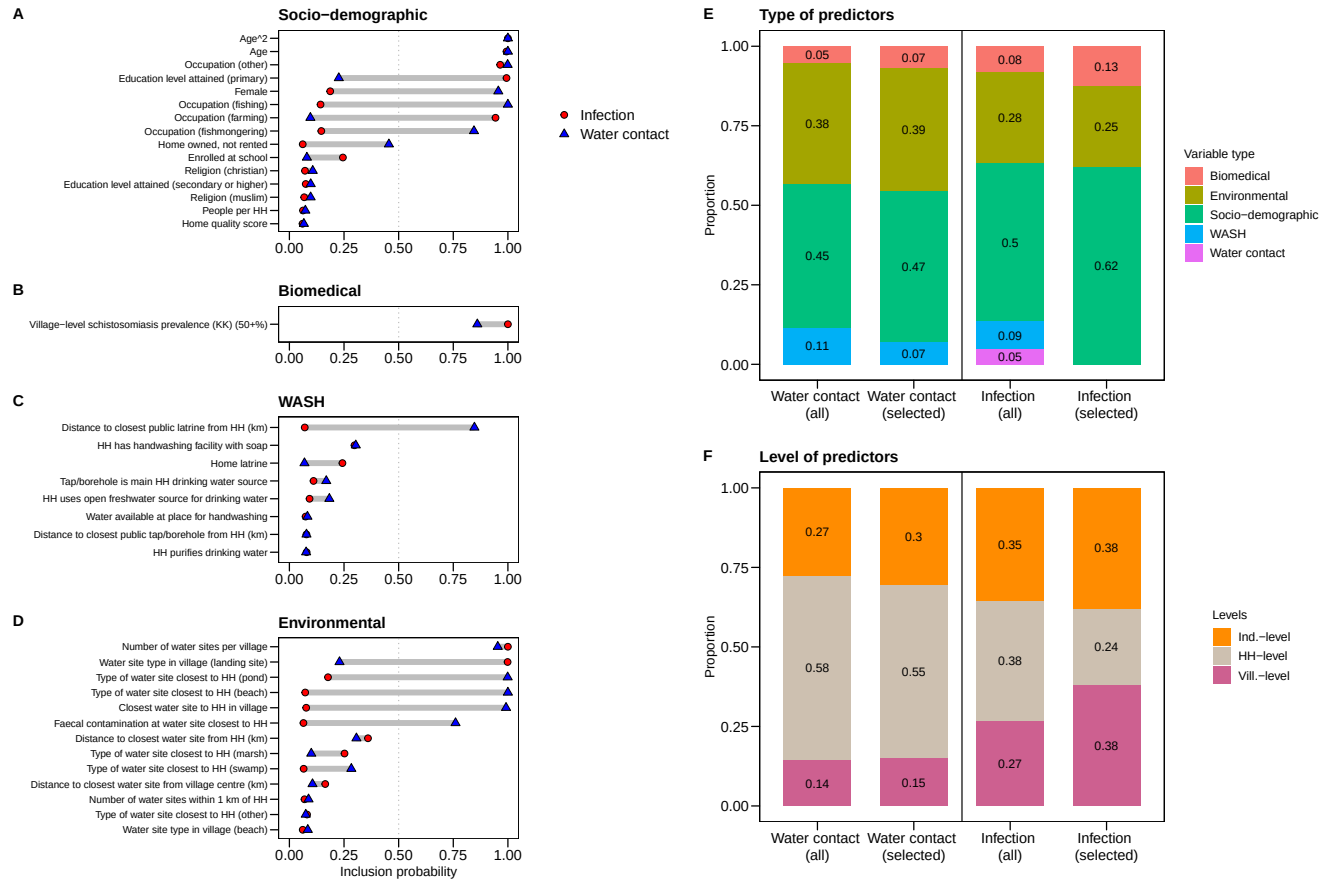


Figure 3.6. Relevant dimensions of water contact and infection. Abbreviations: Prop. = proportion. HH = household. WASH = water, sanitation, and hygiene. Ind.-level = individual-level. Vill.-level = village-level. Panels (A–D) depict the relevance of different types of variables for predicting water contact and infection, as indicated by their inclusion probability in the “best predictive model” (median probability model) based on Bayesian variable selection (BVS). (A) Socio-demographic variables. (B) Biomedical variables. (C) WASH variables. (D) Environmental variables. The dashed vertical lines in (A–D) represent the cut-off ($pr \geq 0.5$) used to select variables for inclusion in the final regression models. Panels (E–F) regroup variables in (A–D) according to their type and level and show the importance of different levels and types of variables as indicated by their proportional representation in the variable set (weighted by their respective inclusion probabilities). (E) Type of predictor variables. (F) Level of predictor variables. Within each box in Panels (E–F), the left column shows the proportion each variable category accounts for within the candidate variable set and the right column shows the importance of each category within the set of selected variables.

3.3.4 Relevant Human-Environmental Dimensions of Water Contact

Bayesian variable selection (BVS) was applied to a comprehensive set of 27 socio-demographic, biomedical, water, sanitation, and hygiene (WASH), and environmental candidate variables (Fig. 3.6a–d). Variables with inclusion probabilities ≥ 0.5 were subsequently used in multivariable logistic regression models to predict water contact. Selected socio-demographic factors included age, age², gender, and occupation. One village-level WASH variable—the distance to the closest public latrine from the household—was included, as well as village infection prevalence. Four environmental variables were selected, including the type of water site closest to the household, the closest water site within the village, the number of water sites per village, and faecal site contamination occurring at the closest site to the household. To summarise which variable types and levels were relevant for predicting water contact, the proportion of variables by group and level was calculated, weighted by their respective inclusion probabilities (Fig. 3.6e–f). Among selected variables, socio-demographics accounted for 47% of all weights, followed by environmental variables at 39%, WASH at 7%, and biomedical variables at 7% (Fig. 3.6e). Individual-level factors accounted for 30%, household-level factors for 55%, and village-level factors for 15% of all weights (Fig. 3.6f).

3.3.5 Determinants of Any Water Contact

Fig. 3.7 presents the main model of any water contact. Significant predictors of current water contact spanned socio-demographic, biomedical, WASH, and environmental variables. Both age (OR 1.18; 95% CI: 1.15–1.21) and age² (OR 0.9978; 95% CI: 0.9975–0.9982) were significantly associated with water contact. The relevance of age² supported a declining relevance of older age with water contact, consistent with the exploratory GAMs. Females had 1.4 times higher odds of having water contact compared to males (95% CI: 1.17–1.68). Fishing and fish mongering occupations were associated with 6.67 times (95% CI: 4.06–10.95), and 2.10 times (95% CI: 1.12–3.95) higher odds of water contact, respectively, when compared to other or no income-earning occupations. Occupation and gender were significantly associated ($\chi^2 = 400.95$, $p < 0.01$). When occupation was removed from the model, the gender coefficient became insignificant (OR 1.05; 95% CI: 0.89–1.24, Fig. S3.4). Observed faecal contamination

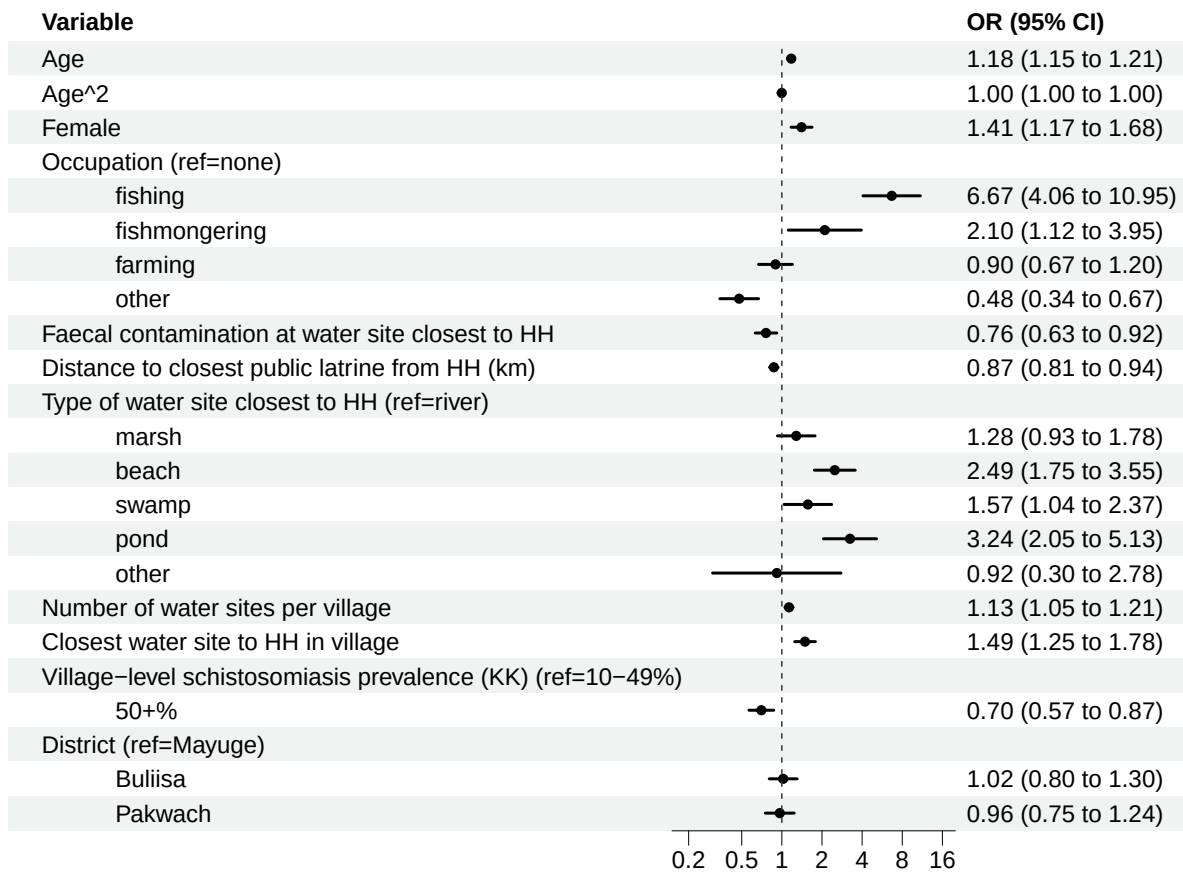


Figure 3.7. Determinants of water contact. Abbreviations: Prop. = proportion. HH = household. KK = Kato–Katz stool microscopy. A logistic regression model is shown for current water contact ($n = 2867$) with 95% confidence intervals from standard errors clustered at the household level. The x-axis is on a log scale. For each categorical variable, we tested group-level significance using likelihood ratio tests (LRTs) comparing the full model to a model without the categorical variable. The global *Moran's I* of the model residuals was 0.064 (95% CI: 0.0639–0.0644, $p < 0.01$).

at the water site closest to the household was associated with 24% lower odds of having water contact (OR 0.76; 95% CI: 0.63–0.92) when compared to living near water sites with no such contamination. Each 1 km-increase in distance to the closest public latrine was associated with 13% lower odds of having water contact (OR 0.87; 95% CI: 0.81–0.94). However, latrine distance was significantly correlated with distance to the closest water site ($\rho_s = 0.33$, $p < 0.01$). Water site ecology influenced water contact behaviour. Individuals who lived closest to a beach, swamp, or pond had significantly higher odds of having water contact compared to people living closest to a river (ORs: 1.57–3.24). Access to water sites, measured by the number of water sites

per village (OR 1.13; 95% CI: 1.05–1.21) and whether the water site closest to the household was within the village (OR 1.49; 95% CI: 1.25–1.78), was positively associated with the odds of water contact. In villages with $\geq 50\%$ *S. mansoni* prevalence, individuals had 30% lower odds of engaging in water contact compared to villages with 10–49% prevalence (OR 0.70; 95% CI: 0.57–0.87). After adjustment for covariates, no significant regional differences in water contact were observed when comparing the Western districts of Pakwach and Buliisa to Mayuge.

Using separate logistic regression models to predict domestic, occupational, and recreational water contact, we found that the effect of gender varied between activity types (Fig. S3.5). Females had 2.43 times higher odds of engaging in domestic water contact than males (OR 2.43; 95% CI: 1.97–3.00), but 72% and 83% lower odds of having recreational and occupational water contact (OR 0.28; 95% CI: 0.11–0.73 and OR 0.17; 95% CI: 0.11–0.27, respectively). Negative binomial models predicting water contact frequency and duration showed no significant gender differences in duration and frequency after adjusting for covariates (Figs. S3.6 and S3.7).

3.3.6 Determinants of Water Contact Versus Infection

We compared determinants of water contact and schistosome infection status, defined as ≥ 1 eggs per gram (EPG) of stool. Overall, 43.3% (1240/2867) of study participants were infected with *S. mansoni*, and 8.2% (236/2867) were heavily infected (≥ 400 EPG). District-level prevalence was highest in Pakwach (50.4%, 477/947), followed by Buliisa (44.1%, 422/958), and Mayuge (35.4%, 341/962). Modelled *S. mansoni* infection prevalence varied by age; 29% (95% CI: 25–33%) of children were infected at age five, prevalence rose to a peak of 63% (95% CI: 59–67%) at age 15, and then declined to 31% (95% CI: 24–36%) at age 50 (Fig. 3.4a). This trend differed from age-dependent water contact, which peaked much later at age 30 and remained at a significantly higher level than infection until age 65. Modelled age-specific infection prevalence trends in GAMs were similar between participants with and without water contact (Fig. S3.8). Infection prevalence did not follow the same linear decline over household distance from water sites, which was observed for water contact (Fig. S3.9). Even crude correlations between current water contact and infection status showed no association ($\rho_s = 0.03$, $p = 0.11$, obs. 2867). Pairwise correlations of infection status ($\rho = 0.16$, $p < 0.01$) or water contact ($\rho_s = 0.08$,

$p < 0.01$) between adults and children within the same household were weak ($n = 2867$). Village-level infection prevalence was uncorrelated with the village-level proportion of individuals with water contact ($\rho_s = 0.004$, $p = 0.98$, $n = 38$).

We selected predictors of infection status using the same methods and candidate set as for water contact (plus six additional water contact variables for the infection model, Table S3.1). The number of variables selected for water contact was ten, compared to seven for infection status. Among these ten variables for water contact, only 50% were selected for inclusion in the infection status model. Shared variables included age, age², occupation, the number of water sites per village, and village infection prevalence. Gender was not selected for the infection model. No water contact or WASH variable met the threshold of inclusion for infection status. Other predictors of water contact not selected for infection included faecal contamination of the water site closest to the household, type of water site closest to the household, closest water site in the village, and distance to the closest public latrine. Two predictors of infection were not selected for the water contact model: education level attained and the landform type of the water site in the village. When grouped by variable type, socio-demographics (accounting for 62% of variables) and environmental variables (25%) were most relevant for infection (Fig. 3.6e). Socio-demographics were more relevant for infection status than for water contact, while environmental variables were less important for infection status compared to water contact. When compared by level, household-level variables were comparatively less relevant for infection status than for water contact, while individual-level variables were more relevant for infection status than for water contact (Fig. 3.6f).

Fig. 3.8 shows results of logistic regressions predicting infection status and heavy infection. Among the five variables consistently selected for water contact and infection, no variable except for age² was significant and of similar magnitude, defined as having overlapping CIs, between the models. Age was positively associated with both outcomes, but the effect size of each one-year increase in age was of greater magnitude for water contact (OR 1.17; 95% CI: 1.15–1.21) than infection (OR 1.05; 95% CI: 1.03–1.08). There was no evidence of association between the two models; no significant correlation was observed between residuals of water contact and infection models ($\rho_s = 0.01$, $p = 0.44$, obs. 2867). Even within infection models, instability

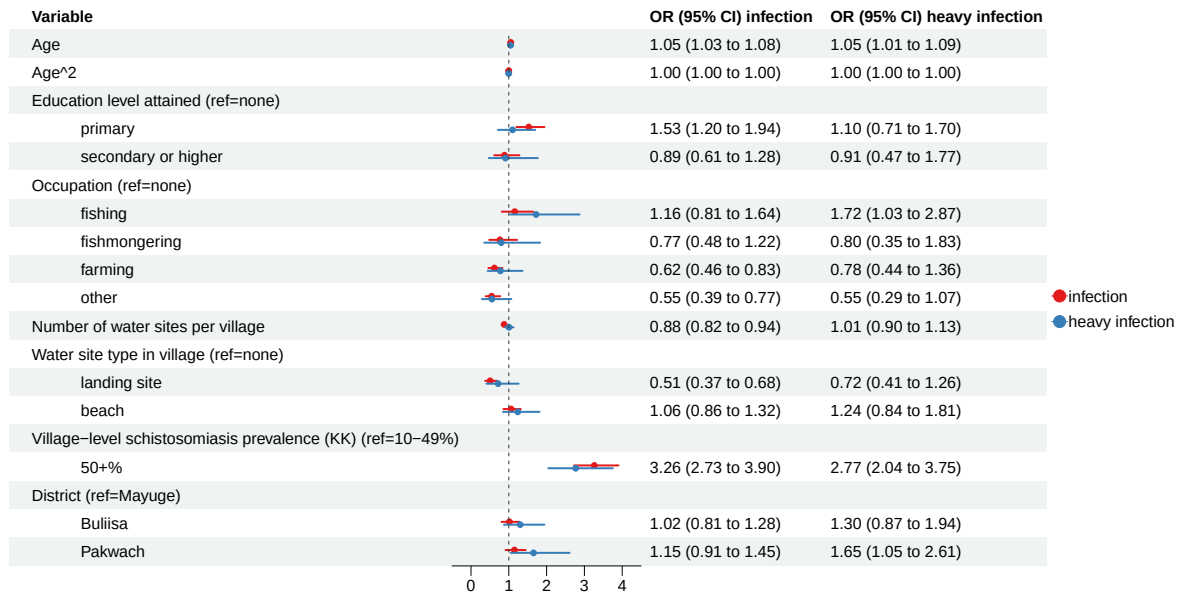


Figure 3.8. Determinants of infection status and heavy infection intensity. Abbreviations: KK = Kato–Katz stool microscopy. Separate logistic regression models predicting *S. mansoni* infection and heavy *S. mansoni* infection ($n = 2867$) with 95% confidence intervals from standard errors clustered at the household level. The x-axis is on a log scale. The global *Moran's I* of the model residuals was 0.0049 (95% CI: 0.0047–0.0052, $p = 0.39$).

was observed across significant variables (Fig. 3.8). Fishing was inconsistently associated with infection, showing only a relationship with heavy infection intensity (400+ EPG) and not infection status (≥ 1 EPG). Having attained primary education was associated with higher odds of infection but not with higher odds of heavy infection. All results remained robust even when infection intensity was not used as a binary variable and run continuously in a negative binomial regression (Fig. S3.10).

3.3.7 Validation of Water Contact and Infection Measurements

We investigated water contact misclassification in our survey data by investigating patterns in direct water contact observations. Directly observed water contacts at the most used sites within study villages were available for approximately 32% (12/38) of study villages and were collected at the same time as the self-reported data. For the age structure of water contacts, the self-reported water contact data closely resembled observed community-wide water contact patterns from direct observation (Fig. S3.11). To address the possibility of missed light infections due to the low sensitivity of Kato–Katz microscopy, negative Kato–Katz results were recoded as

positive when POC-CCA diagnostic results were positive, where trace was classified as negative. Infection results remained similar after reclassification of infection status (Fig. S3.12).

3.3.8 Predictive Capacity of Water Contact Models by Variable Selection Technique

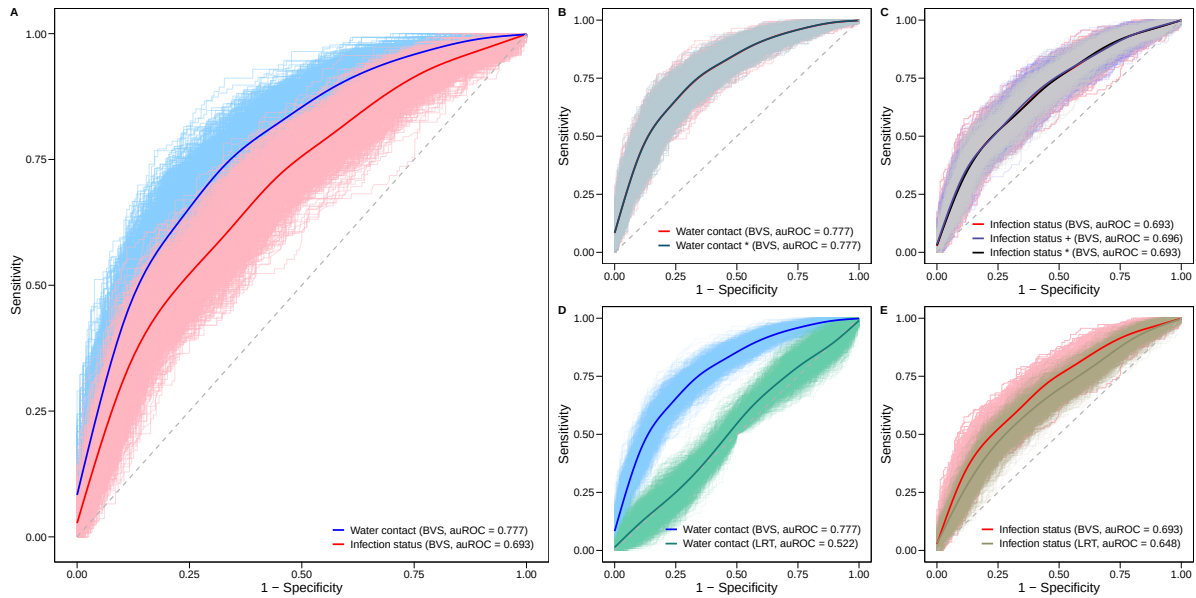


Figure 3.9. Predictive capacity of water contact and infection models. (A) Comparison of auROC for the water contact versus infection status from models in Figs. 3.7 and 3.8 with variables selected using BVS. (B) Comparison of the main water contact model (Fig. 3.7) versus a more granular water contact model allowing for the selection of additional variables from a set of seven snail variables, water contact* model, selected using BVS. (C) Comparison of auROC for the infection status model in Fig. 3.8 versus auROC of two more granular infection status models, the infection⁺ model and the infection* model (Figs. S3.15 and S3.16). In the infection⁺ model variable selection process, we allowed for the selection of additional variables from a set of 82 granular water contact variables using BVS. In the infection* model variable selection process, we allowed for the selection of additional variables from a set of seven snail variables using BVS. (D) Comparison of auROC for the water contact model selected using BVS versus a water contact model with variables selected using likelihood ratio tests (LRTs). LRTs compared one variable at a time against an “empty” model with village-level fixed effects at $p < 0.05$ to select variables from the same candidate set as BVS. Differences between variable sets selected via BVS and LRTs are shown in Fig. S3.14. (E) Comparison of auROC for the infection status model using BVS versus an infection model with variables selected using LRTs. All auROCs were computed using 100 repeats of ten-fold cross-validation.

The predictive performance of the water contact and infection models was assessed using ten-fold cross-validation, quantifying the influence of variable selection method (BVS versus LRTs) and the relevance of including additional snail and water contact variables (Fig. 3.9). The water contact model had better predictive performance than the infection status model, with an

area under the receiver operating curve (auROC) of 0.777 versus 0.693 ($p < 0.01$, Fig. 3.9a). Neither the inclusion of additional snail variables nor the inclusion of more granular water contact variables substantially improved the predictive accuracy of any model (Fig. 3.9b–c). BVS outperformed LRTs for water contact prediction (auROC: 0.777 versus 0.521, respectively, $p < 0.01$, Fig. 3.9d, Fig. S3.10). Compared to water contact, variable sets selected for infection status were more comparable between BVS and LRTs (Table S3.6), and the gap in performance was smaller (auROC: 0.696 and 0.648, $p < 0.01$, Fig. 3.9e). There was weak residual spatial correlation in the water contact model (*Moran's I* = 0.064, $p < 0.01$).

3.4 Discussion

The 2022 WHO guideline on schistosomiasis control and elimination as well as recent cross-NTD guidelines emphasise the need to complement MDA with behaviour change and WASH to achieve and sustain reductions in schistosome infection prevalence [17, 43]. However, this requires an in-depth understanding of the determinants of water contact. In this chapter, a data-driven, population-based analysis was conducted of 2867 individuals within a community-based cohort, SchistoTrack [12], in 38 diverse villages in three districts in Uganda. This analysis characterised complex water contact and *S. mansoni* infection patterns over age and gender. We developed statistical models and a generalisable analysis pipeline for the wider research community to untangle the intricate dynamics of water contact and infection. This analysis included a comprehensive set of exposure variables. We focused on current water contact using detailed individual, household, and environmental predictors of water contact as well as WASH variables. We found that current water contact was not strongly associated with current schistosome infection, highlighting the complex relationship between exposure and infection. The findings should not be misconstrued as implying that water contact is irrelevant for infection, as all infections are caused by contact at some point with infested freshwater sources.

There was no significant village-level correlation between the proportion of individuals with water contact and village-level infection prevalence. Hotspots exist in Uganda and elsewhere and are commonly defined as high prevalence areas ($\geq 50\%$ by Kato–Katz microscopy) that are

unresponsive to repeated MDA [44, 45]. SchistoTrack was conducted in villages with $\geq 10\%$ infection prevalence in districts that each had received over 13 annual rounds of MDA since the year 2003 [14]. Individuals in hotspot villages ($\geq 50\%$ infection prevalence) had lower odds of water contact when compared to villages with 10–49% prevalence. There have been calls for intensified control in hotspot settings, including the use of exposure reduction interventions [46]. Our findings suggest that such intensified measures, if intended to reduce water contact, should be targeted not only in hotspots but also in moderate prevalence settings.

Age-specific infection patterns did not approximate age-specific water contact, suggesting an important role of acquired immunity in explaining infection patterns over age. Despite our population being from three districts with varying local climates, geography, and tribal and religious groups, water contact was comparatively low in children across all districts. Infection prevalence peaked at age 15—more than a decade earlier than water contact, which peaked at age 30. Our finding is consistent with cross-sectional studies conducted before widespread MDA in endemic settings that reported lower levels of reinfection in adults despite higher levels of water contact [47, 48]. The discrepant age trends of water contact and infection support the slow build-up of immunity, possibly due to immunity acquired through repeated exposure to antigens from the death of mature flukes, which may reduce susceptibility to reinfection [49, 50]. Given low levels of water contact among children compared to adults, we speculate that even longitudinal water contact data would be unlikely to reveal much more substantive correlations between measures of cumulative water contact and infection. This chapter suggests that the relationship between water contact (as an element of exposure) and acquired immunity associated with water contact is likely highly non-linear; the potential critical threshold of exposure required to explain the sudden drop in infection levels at age 15 years would be suspected to be incredibly low. In endemic settings, children typically become infected early in life, with one study in Kenya showing that the seroprevalence based on antibody response to *S. mansoni* soluble egg antigen was 38% at age two and 63% at age five [8]. For immunologically naïve children, low levels of water contact are likely sufficient for them to become infected. Due to repeated exposure, children acquire a higher burden of infection over time and slowly start to mount a protective immune response, which likely downregulates their infection intensity despite high levels of

exposure [51, 52]. Understanding the temporal dynamics of exposure, infection, and immunity is an important topic for future research. In this analysis, it was not possible to rule out the role of the timing of first infection, longevity of single infections, influence of co-infections and other immune-modulating factors, or age-specific immune responses [6, 8]. There is a need for future longitudinal studies that combine the collection of exposure histories with immunological data. In terms of modelling, this chapter advocates for relaxing assumptions about a direct correspondence between current age-specific water contact rates and infection prevalence.

We observed clear gender differences in the dominant types of water contact among adults. Female adults mostly had domestic water contact, while male adults primarily engaged in occupational water contact. These activities typically take place at different times of the day and involve varying levels of bodily immersion, which may lead to differing exposure levels. A systematic review suggested that gender differences in infection prevalence might be explained by these varied exposure patterns [53]. While the review was unable to account for age differences due to a lack of age-disaggregated reporting, the results here emphasise the need for age stratification. We found that water contact patterns were similar in males and females up to age 15. Among adults, being female was positively associated with water contact, but only after adjusting for occupation, underscoring the interrelatedness of gender and occupation. When using a composite outcome, such as any water contact, it remains challenging to isolate the effects of gender. However, by separately modelling different water contact activities, we found that females had significantly higher domestic water contact but significantly lower occupational and recreational water contact. To promote gender equity in exposure reduction programs, both domestic and occupational water contact should be addressed, as they represent the primary sources of exposure for females and males, respectively.

Past research has tried to separate environmental and human-behavioural elements of exposure [54, 55], yet these factors are intricately linked. Here, environmental features influenced both water contact behaviour and environmental hazards, where hazard is defined as the local abundance of schistosome cercariae. We identified several environmental factors, such as the number of water sites within the village, site type, and faecal contamination, which impacted

water contact patterns. Some of these factors, especially faecal contamination, also impact environmental hazard. When we included several snail-related factors, such as the number of snails or infected snails at the nearest site, we found that they did not influence water contact behaviour as would be expected given that they cannot be easily observed by individuals. Understanding which ecological factors influence water contact, environmental hazard, or both, can shed light on the complex human-environmental processes involved in exposure and clarify how environmental factors should be most appropriately integrated into exposure indices and parameterised in models. Characterising how environmental risk varies between different types of water settings, such as rivers, swamps, and lakes, is important for quantifying the risk associated with different water bodies and targeting treatments and environmental control interventions accordingly.

Distance to waterbodies has been used for approximating the force of infection, sampling individuals for infection studies, and choosing where to implement MDA [56, 57]. Many studies have used binary indicators of distance with varying cut-offs ranging from 0.5–5 km to predict infection [2, 3]. This chapter found a clear, almost linear gradient of the distance of a home to freshwater bodies with water contact. The proportion of participants with water contact declined by 34% as household distance increased from 0–1 km to the closest water site. Infection prevalence showed a weaker decline of only 19% from 0–1 km for household distance to water sites, which appeared non-linear with a much less discernible trend than water contact. In our statistical models, waterbody distance was not selected because of the inclusion of occupation, water site availability, and other distance-related factors that were able to account for spatial patterns in water contact and infection, as indicated by low residual spatial autocorrelation. Different distance measures have been used interchangeably elsewhere, such as household distance [9], village distance [10], or school distance [19]. We showed that these measures were not equivalent, as we found no water contact gradients in school-based measures. Thus, only household or village distance to large freshwater bodies should be used to estimate exposure gradients. And, critically, whether to use the proportion of individuals with current water contact or infection prevalence as outcomes for estimating distance gradients should be carefully considered in future studies and interventions dependent on the use case.

3.4.1 Limitations

A key challenge for exposure studies has been reliable water contact measurement. Self-reported measures have been the most widely used [58] but are considered less reliable than direct observations or wearable GPS logger data [13, 59]. Here, we validated the use of self-reported water contact data by comparing it with direct observations, demonstrating that surveys provide a cost-effective way to obtain valid population-level water contact measures. We note however, that while we found a good representation of overall community-level trends in self-reported data, some behaviours such as recreational water contact of children could be under-reported as water contact was reported by the household head. It remains possible that children are sent to retrieve drinking water, hence reporting this activity, but also go for a leisurely swim which was not reported. There remains a need for future water contact ascertainment studies that match individuals directly with more granular measurements, and when such ascertainment is completed, we suggest rerunning our models with regression dilution to investigate the influence of possible measurement error. For future research, more granular measurements of frequency and duration of water contact should be investigated using wearable GPS loggers.

3.4.2 Conclusion

Due to the complexity of water contact, we suggest that appropriate measurement of water contact should be tailored to the specific goal of a national programme or research study. For instance, for monitoring and surveillance surveys, especially when conducted alongside MDA, collecting age, gender, and occupation status may be sufficient to ascertain basic risk factors for water contact. Home location data could be collected to delineate a buffer area in which individuals are at high exposure risk. By contrast, sentinel surveillance sites based at schools are unlikely to identify fine-scale gradients in water contact. For interventions, data on common water contact activities and their typical time of day could aid in identifying target groups and behaviours. With finer spatio-temporal measurements that might include indicators from wearable GPS loggers, in the future, it will be possible to identify the exact durations and intensity of water contact, seasonal and diurnal water contact patterns, and explore the role of human mobility for schistosome transmission.

This chapter found that human-environmental interactions were highly group-specific, suggesting that different demographic groups need to be targeted with specific (WASH) interventions. Domestic activities were the dominant type of water contact for children and women. Additional research is needed to evaluate whether targeted safe water infrastructure provision interventions can successfully redirect contact due to laundry and getting drinking water to safe sources. Despite the inclusion of detailed household-level WASH variables, only distance to public latrines was significantly associated with water contact. Surprisingly, the use of safe water sources and household water purification were not relevant for water contact. Given that water contact was dependent on occupation for men, there was no clear suggestion that WASH would influence water contact for this group. Future studies might instead focus on providing protective equipment such as gum boots or gloves for fishermen, building safe landing sites for fishing boats, or on socio-economic interventions such as cash supplements or employment retraining, to understand how to reduce water contact for men. This analysis focused on occupational fishing as opposed to recreational/irregular fishing. If a more comprehensive view of any fishing was undertaken, it may be possible for future research to identify interventions that may shift fishing activities to less risky times of day, to modify the predominant type of fishing activity, or to test the impact of protective equipment such as gumboots or of building safe landing sites.

The influence of distance to waterbodies on water contact behaviour is likely mediated by human mobility. People of different ages and genders may have varying typical ranges of movement, which influence how far they typically travel to access water sites. However, as the data used here were self-reported and did not include specific site information, we were unable to capture which specific water sites individuals used. Future work should use data from mobile phones, GPS loggers, or other location tracking devices to understand the geographic extent of mobility, which is crucial for designing control and elimination programmes. Such mobility information would help to define which specific geographic zones/units interventions will need to cover.

The results of this chapter provide a compendium of water contact metrics and patterns to facilitate multifaceted control of schistosomiasis. In SchistoTrack, children were highly

susceptible to infection despite relatively low levels of water contact, providing support for continued MDA. Given the important role of adults in community-level water contact, exposure reduction interventions to reduce water contact should be targeted at all age groups even in cases where MDA is limited to school-age children. Our study supplements the current WHO guideline for schistosomiasis treatment [17], which defines at-risk communities and groups based on infection prevalence and intensity, by providing profiling priority groups with high water contact. Concerning the guidelines for selecting implementation units for MDA, water contact was heavily concentrated in households within 0.43 km of water sites, while infection prevalence remained high beyond 1 km distance, suggesting that water contact interventions should be implemented at a smaller spatial scale than MDA. With detailed data representing diverse geographies in Uganda and high predictive model performance, the findings have high internal validity. To further assess the generalisability of the findings presented in this chapter, the same methods should be applied to data from different countries and different schistosome species.

This chapter highlights the complexity of human exposure and the need for guidelines to consider water contact when determining the frequency of treatment. It also provides information for the WHO to create measurement tools to enable standardised data collection on water contact for use in country monitoring and control programmes.

References

1. Scott J, Diakhate M, Vereecken K, Fall A, Diop M, Ly A, De Clercq D, Vlas S de, Berkvens D, Kestens L, and Gryseels B. Human water contacts patterns in *Schistosoma mansoni* epidemic foci in northern Senegal change according to age, sex and place of residence, but are not related to intensity of infection. *Tropical Medicine & International Health* 2003 Feb; 8:100–8. DOI: 10.1046/j.1365-3156.2003.00993.x
2. Rudge JW, Stothard JR, Basáñez MG, Mgeni AF, Khamis IS, Khamis AN, and Rollinson D. Micro-epidemiology of urinary schistosomiasis in Zanzibar: Local risk factors associated with distribution of infections among schoolchildren and relevance for control. *Acta Tropica* 2008 Jan; 105:45–54. DOI: 10.1016/j.actatropica.2007.09.006
3. Lund AJ, Sokolow SH, Jones IJ, Wood CL, Ali S, Chamberlin A, Sy AB, Sam MM, Jouanard N, Schacht AM, Senghor S, Fall A, Ndione R, Riveau G, De Leo GA, and Lopez-Carr D. Exposure, hazard, and vulnerability and their contribution to *Schistosoma haematobium* re-infection in northern Senegal. *The Lancet Planetary Health* 2021 Apr; 5:S10. DOI: 10.1016/S2542-5196(21)00094-2
4. Reitzug F, Ledien J, and Chami GF. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2023 Jun; 17:e0011377. DOI: 10.1371/journal.pntd.0011377
5. Pearce EJ and MacDonald AS. The immunobiology of schistosomiasis. *Nature Reviews Immunology* 2002 Jul; 2:499–511. DOI: 10.1038/nri843

6. Joseph S, Jones FM, Walter K, Fulford AJ, Kimani G, Mwatha JK, Kamau T, Kariuki HC, Kazibwe F, Tukahebwa E, Kabatereine NB, Ouma JH, Vennervald BJ, and Dunne DW. Increases in human T helper 2 cytokine responses to *Schistosoma mansoni* worm and worm-tegument antigens are induced by treatment with praziquantel. *The Journal of Infectious Diseases* 2004 Aug; 190:835–42. doi: 10.1086/422604
7. Cytokine production in whole blood cultures from a fishing community in an area of high endemicity for *Schistosoma mansoni* in Uganda: the differential effect of parasite worm and egg antigens. 72. doi: 10.1128/IAI.72.2.728–734.2004
8. Arnold BF, Kanyi H, Njenga SM, Rawago FO, Priest JW, Secor WE, Lammie PJ, Won KY, and Odiere MR. Fine-scale heterogeneity in *Schistosoma mansoni* force of infection measured through antibody response. *Proceedings of the National Academy of Sciences* 2020 Aug; 117:23174–81. doi: 10.1073/pnas.2008951117
9. Lamberti O, Kabatereine NB, Tukahebwa EM, and Chami GF. *Schistosoma mansoni* infection risk for school-aged children clusters within households and is modified by distance to freshwater bodies. *PLoS ONE* 2021 Nov; 16:e0258915. doi: 10.1371/journal.pone.0258915
10. Anderson R, Turner H, Farrell S, and Truscott J. Studies of the Transmission Dynamics, Mathematical Model Development and the Control of Schistosome Parasites by Mass Drug Administration in Human Communities. *Mathematical Models for Neglected Tropical Diseases - Essential Tools for Control and Elimination, Part B*. Elsevier, 2016 :199–246. doi: 10.1016/bs.apar.2016.06.003
11. Graham M, Ayabina D, Lucas TC, Collyer BS, Medley GF, Hollingsworth TD, and Toor J. SCHISTOX: An individual based model for the epidemiology and control of schistosomiasis. *Infectious Disease Modelling* 2021; 6:438–47. doi: 10.1016/j.idm.2021.01.010
12. Nuffield Department of Population Health, University of Oxford. SchistoTrack: a prospective multimorbidity cohort. <https://www.bdi.ox.ac.uk/research/schistotrack>. [Accessed on: 2024 Apr 16]
13. Trienekens SCM, Faust CL, Besigye F, Pickering L, Tukahebwa EM, Seeley J, and Lamberton PHL. Variation in water contact behaviour and risk of *Schistosoma mansoni* (re)infection among Ugandan school-aged children in an area with persistent high endemicity. *Parasites & Vectors* 2022 Dec; 15:15. doi: 10.1186/s13071-021-05121-6
14. Anjorin S, Nabatte B, Mpooya S, Tinkitina B, Opio CK, Kabatereine NB, and Chami GF. Epidemiology of periportal fibrosis and relevance of current *Schistosoma mansoni* infection within the context of repeated mass drug administration in rural Uganda: a population-based, cross-sectional study. *The Lancet Microbe* 2024 Dec; 5:100944. doi: 10.1016/j.lanmic.2024.07.007
15. Piantadosi S, Byar DP, and Green SB. The ecological fallacy. *American journal of epidemiology* 1988; 127:893–904
16. Reitzug F, Kabatereine NB, Byaruhanga AM, Besigye F, Nabatte B, and Chami GF. Current water contact and *Schistosoma mansoni* infection have distinct determinants: a data-driven population-based study in rural Uganda. *Nature Communications* 2024 Nov; 15. doi: 10.1038/s41467-024-53519-4
17. World Health Organization. WHO guideline on control and elimination of human schistosomiasis. Geneva: World Health Organization, 2022
18. Phillips AE, Gazzinelli-Guimaraes PH, Aurelio HO, Ferro J, Nala R, Clements M, King CH, Fenwick A, Fleming FM, and Dhanani N. Assessing the benefits of five years of different approaches to treatment of urogenital schistosomiasis: A SCORE project in Northern Mozambique. *PLoS Neglected Tropical Diseases* 2017 Dec; 11:e0006061. doi: 10.1371/journal.pntd.0006061
19. Musuva R, Odiere M, Mwinzi P, Omondi I, Rawago F, Matendechero S, Kittur N, Campbell C, and Colley D. Unprotected water sources and low latrine coverage are contributing factors to persistent hotspots for schistosomiasis in western Kenya. *PLoS ONE* 2021 Sep; 16. doi: 10.1371/journal.pone.0253115
20. Assaré RK, N'Tamon RN, Bellai LG, Koffi JA, Mathieu TBI, Ouattara M, Hürlimann E, Coulibaly JT, Diabaté S, N'Goran EK, and Utzinger J. Characteristics of persistent hotspots of *Schistosoma mansoni* in western Côte d'Ivoire. *Parasites & Vectors* 2020 Jul; 13. doi: 10.1186/s13071-020-04188-x
21. Graeff-Teixeira C, Favero V, Pascoal VF, De Souza RP, Rigo FDV, Agnese LHD, Bezerra FSM, Coelho PMZ, Enk MJ, Favre TC, Katz N, Oliveira RR, Dos Reis MG, and Pieri OS. Low specificity of point-of-care circulating cathodic antigen (POC CCA) diagnostic test in a non-endemic area for schistosomiasis mansoni in Brazil. *Acta Tropica* 2021 May; 217:105863. doi: 10.1016/j.actatropica.2021.105863

22. Peralta JM and Cavalcanti MG. Is POC-CCA a truly reliable test for schistosomiasis diagnosis in low endemic areas? The trace results controversy. *PLoS Neglected Tropical Diseases* 2018 Nov; 12:e0006813. doi: 10.1371/journal.pntd.0006813
23. Colley DG, Andros TS, and Campbell CH. Schistosomiasis is more prevalent than previously thought: what does it mean for public health goals, policies, strategies, guidelines and intervention programs? *Infectious Diseases of Poverty* 2017 Mar; 6. doi: 10.1186/s40249-017-0275-5
24. Pinot de Moira A, Fulford AJC, Kabatereine NB, Ouma JH, Booth M, and Dunne DW. Analysis of Complex Patterns of Human Exposure and Immunity to *Schistosomiasis mansoni*: The Influence of Age, Sex, Ethnicity and IgE. *PLoS Neglected Tropical Diseases* 2010 Sep; 4:e820. doi: 10.1371/journal.pntd.0000820
25. Altman DG and Royston P. The cost of dichotomising continuous variables. *BMJ* 2006 May; 332:1080.1. doi: 10.1136/bmj.332.7549.1080. [Accessed on: 2024 Apr 16]
26. Chami GF, Kontoleon AA, Bulte E, Fenwick A, Kabatereine NB, Tukahebwa EM, and Dunne DW. Profiling Nonrecipients of Mass Drug Administration for Schistosomiasis and Hookworm Infections: A Comprehensive Analysis of Praziquantel and Albendazole Coverage in Community-Directed Treatment in Uganda. *Clinical Infectious Diseases* 2016 Sep; 62:200–7. doi: 10.1093/cid/civ829
27. Grimes JET, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The Relationship between Water, Sanitation and Schistosomiasis: A Systematic Review and Meta-analysis. *PLoS Neglected Tropical Diseases* 2014 Dec; 8:e3296. doi: 10.1371/journal.pntd.0003296
28. World Health Organization and UNICEF. Safely managed drinking water - thematic report on drinking water 2017. Tech. rep. Geneva: World Health Organization and UNICEF, 2017
29. Miguel E and Kremer M. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 2004 Jan; 72:159–217. doi: 10.1111/j.1468-0262.2004.00481.x
30. Institute WR. Waterbodies in Uganda. <https://datasets.wri.org/dataset/waterbodies-in-uganda>. 2010
31. Wood S. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. <https://cran.r-project.org/web/packages/mgcv/index.html>
32. Clyde M. BAS: Bayesian variable selection and model averaging using Bayesian adaptive sampling. <https://cran.r-project.org/web/packages/BAS/index.html>
33. Clyde MA, Ghosh J, and Littman ML. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics* 2011 Jan; 20:80–101. doi: 10.1198/jcgs.2010.09049
34. Liang F, Paulo R, Molina G, Clyde MA, and Berger JO. Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association* 2008 Mar; 103:410–23. doi: 10.1198/016214507000001337
35. Berger JO, Pericchi LR, Ghosh J, Samanta T, De Santis F, Berger J, and Pericchi L. Objective Bayesian methods for model selection: introduction and comparison. *Lecture Notes-Monograph Series* 2001 :135–207
36. Bayarri M and Garcia-Donato G. Extending conventional priors for testing general hypotheses in linear models. *Biometrika* 2007 Feb; 94:135–52. doi: 10.1093/biomet/asm014
37. Barbieri MM and Berger JO. Optimal predictive model selection. *The Annals of Statistics* 2004 Jun; 32. doi: 10.1214/009053604000000238
38. Zeileis A, Köll S, and Graham N. Various versatile variances: an object-oriented implementation of clustered covariances in R. *Journal of Statistical Software* 2020; 95. doi: 10.18637/jss.v095.i01
39. Hausman JA. Specification tests in econometrics. *Econometrica* 1978 Nov; 46:1251. doi: 10.2307/1913827. [Accessed on: 2024 Aug 8]
40. Lüdtke D, Makowski D, Ben-Shachar MS, Patil I, Waggoner P, Wiernik BM, and Thériault R. performance: assessment of regression models performance. <https://CRAN.R-project.org/package=performance>. 2019 Apr. [Accessed on: 2024 Aug 8]
41. Fox J and Monette G. Generalized collinearity diagnostics. *Journal of the American Statistical Association* 1992 Mar; 87:178–83. doi: 10.1080/01621459.1992.10475190
42. Wolmarans C, Kock K de, Strauss H, and Bornman M. Daily emergence of *Schistosoma mansoni* and *S. haematobium* cercariae from naturally infected snails under field conditions. *Journal of Helminthology* 2002 Sep; 76:273–7. doi: 10.1079/joh2002122

43. World Health Organization. WASH and Health Working Together: A ‘how-to’ Guide for Neglected Tropical Disease Programmes. World Health Organization, 2023
44. Wiegand RE, Mwinzi PNM, Montgomery SP, Chan YL, Andiego K, Omedo M, Muchiri G, Ogutu MO, Rawago F, Odiere MR, Karanja DMS, and Secor WE. A Persistent Hotspot of *Schistosoma mansoni* Infection in a Five-Year Randomized Trial of Praziquantel Preventative Chemotherapy Strategies. *The Journal of Infectious Diseases* 2017 Sep; 216:1425–33. doi: 10.1093/infdis/jix496
45. Kittur N, Campbell CH, Binder S, Shen Y, Wiegand RE, Mwanga JR, Kinung’hi SM, Musuva RM, Odiere MR, Matendehero SH, Knopp S, and Colley DG. Discovering, defining, and summarizing persistent hotspots in SCORE studies. *The American Journal of Tropical Medicine and Hygiene* 2020 Jul; 103:24–9. doi: 10.4269/ajtmh.19-0815. [Accessed on: 2024 Apr 16]
46. Amoah AS, Hoekstra PT, Casacuberta-Partal M, Coffeng LE, Corstjens PLAM, Greco B, Lieshout L van, Lim MD, Markwalter CF, Odiere MR, Reinhard-Rupp J, Roestenberg M, Stothard R, Tchuem Tchuente LA, Vlas SJ de, and Dam GJ van. Sensitive diagnostic tools and targeted drug administration strategies are needed to eliminate schistosomiasis. *The Lancet Infectious Diseases* 2020 Jul; 20:e165–e172. doi: 10.1016/S1473-3099(20)30254-1
47. Wilkins H, Blumenthal U, Hagan P, Hayes R, and Tulloch S. Resistance to reinfection after treatment of urinary schistosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1987 Jan; 81:29–35. doi: 10.1016/0035-9203(87)90273-2
48. Kabatereine NB, Vennervald BJ, Ouma JH, Kemijumbi J, Butterworth AE, Dunne DW, and Fulford AJC. Adult resistance to schistosomiasis mansoni: age-dependence of reinfection remains constant in communities with diverse exposure patterns. *Parasitology* 1999 Jan; 118:101–5. doi: 10.1017/S0031182098003576
49. Fitzsimmons CM, Jones FM, Pinot de Moira A, Protasio AV, Khalife J, Dickinson HA, Tukahebwa EM, and Dunne DW. Progressive Cross-Reactivity in IgE Responses: an Explanation for the Slow Development of Human Immunity to Schistosomiasis? *Infection and Immunity* 2012 Dec; 80:4264–70. doi: 10.1128/iai.00641-12
50. Woolhouse M and Hagan P. Seeking the ghost of worms past. *Nature Medicine* 1999 Nov; 5:1225–7. doi: 10.1038/15169
51. Hagan P, Blumenthal UJ, Dunn D, Simpson AJG, and Wilkins HA. Human IgE, IgG4 and resistance to reinfection with *Schistosoma haematobium*. *Nature* 1991 Jan; 349:243–5. doi: 10.1038/349243a0
52. Woolhouse M. Patterns in Parasite Epidemiology: The Peak Shift. *Parasitology Today* 1998 Oct; 14:428–34. doi: 10.1016/S0169-4758(98)01318-0
53. Ayabina DV, Clark J, Bayley H, Lamberton PHL, Toor J, and Hollingsworth TD. Gender-related differences in prevalence, intensity and associated risk factors of *Schistosoma* infections in Africa: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2021 Nov; 15:e0009083. doi: 10.1371/journal.pntd.0009083
54. Jones IJ, Sokolow SH, Chamberlin AJ, Lund AJ, Jouanard N, Bandagny L, Ndione R, Senghor S, Schacht AM, Riveau G, Hopkins SR, Rohr JR, Remais JV, Lafferty KD, Kuris AM, Wood CL, and De Leo G. Schistosome infection in Senegal is associated with different spatial extents of risk and ecological drivers for *Schistosoma haematobium* and *S. mansoni*. *PLoS Neglected Tropical Diseases* 2021 Sep; 15:e0009712. doi: 10.1371/journal.pntd.0009712
55. Wood CL, Sokolow SH, Jones IJ, Chamberlin AJ, Lafferty KD, Kuris AM, Jocque M, Hopkins S, Adams G, Buck JC, Lund AJ, Garcia-Vedrenne AE, Fiorenza E, Rohr JR, Allan F, Webster B, Rabone M, Webster JP, Bandagny L, Ndione R, Senghor S, Schacht AM, Jouanard N, Riveau G, and De Leo GA. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. *Proceedings of the National Academy of Sciences* 2019 Oct; 116:23182–91. doi: 10.1073/pnas.1903698116
56. Booth M, Vennervald BJ, Kenty L, Butterworth AE, Kariuki HC, Kadzo H, Ileri E, Amaganga C, Kimani G, Mwatha JK, Otedo A, Ouma JH, Muchiri E, and Dunne DW. Micro-geographical variation in exposure to *Schistosoma mansoni* and malaria, and exacerbation of splenomegaly in Kenyan school-aged children. *BMC Infectious Diseases* 2004 May; 4. doi: 10.1186/1471-2334-4-13
57. Seto EYW, Lee YJ, Liang S, and Zhong B. Individual and village-level study of water contact patterns and *Schistosoma japonicum* infection in mountainous rural China. *Tropical Medicine & International Health* 2007 Oct; 12:1199–209. doi: 10.1111/j.1365-3156.2007.01903.x

CHAPTER 3. WATER CONTACT SURVEY DATA ANALYSIS

58. Bethony J, Williams JT, Brooker S, Gazzinelli A, Gazzinelli MF, LoVerde PT, Corrêa-Oliveira R, and Kloos H. Exposure to *Schistosoma mansoni* infection in a rural area in Brazil. Part III: household aggregation of water-contact behaviour. *Tropical Medicine & International Health* 2004 Feb; 9:381–9. doi: 10.1111/j.1365-3156.2004.01203.x
59. Seto EYW, Sousa-Figueiredo JC, Betson M, Byalero C, Kabatereine NB, and Stothard JR. 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. *Geospatial health* 2012 Nov; 7:1. doi: 10.4081/gh.2012.99

3.5 Supplementary Information

Supplementary Tables

Table S3.1. Variable definitions

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Age	Age to the nearest year	Continuous		Socio-demographic	Ind.-level	X	X
Age ²	Age to the nearest year squared	Continuous		Socio-demographic	Ind.-level	X	X
Female	Self-reported gender	Binary		Socio-demographic	Ind.-level	X	X
Education level attained	Highest level of education completed	Categorical	none, primary, secondary or higher	Socio-demographic	Ind.-level	X	X
Religion	Religion	Categorical	no religion/other, christian, muslim	Socio-demographic	Ind.-level	X	X
Enrolled at school	Participant currently enrolled (all participants aged 18+ coded as not enrolled)	Binary		Socio-demographic	Ind.-level	X	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Domestic water contact	Participant engages in one or more of the following domestic water contact activities: getting drinking water, washing clothes with soap, washing clothes without soap, washing jerry cans or household items, washing clothes with soap	Binary		Water contact	Ind.-level	-	X
Occupational water contact	Participant engages in one or more of the following occupational water contact activities: collecting papyrus, fishing, fishmongering, collecting shells	Binary		Water contact	Ind.-level	-	X
Recreational water contact	Participant engages in one or more of the following recreational water contact activities: swimming or playing	Binary		Water contact	Ind.-level	-	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Water contact frequency per week	Water contact frequency per individual per week across all activities (generated by combining the frequency of all individual activities)	Continuous		Water contact	Ind.-level	-	X
Water contact duration per week (hrs)	Water contact duration in hours per week across all activities (generated by combining frequency and duration of all individual activities)	Categorical		Water contact	Ind.-level	-	X
Number of water contact activities	Number of distinct water contact activities a participant engaged in	Discrete		Water contact	Ind.-level	-	X
Occupation	Occupation of the household head (aged 18+), for all participants aged ≥18 occupation coded as 'none'	Categorical	none, fishing, fishmongering, farming, other	Socio-demographic	HH-level	X	X
Home owned, not rented	Household owns home, is not renting	Binary		Socio-demographic	HH-level	X	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Home quality score	Quality of the roof, wall, and floor ranked from 1-4 and summed. The rank order of the materials was grass, sticks, plastic, and metal for the roof; mud and sticks, plastic, metal, and bricks or cement for the walls; mud, plastic, wood planks, and brick or cement for the floor.	Discrete		Socio-demographic	HH-level	X	X
People per HH	Number of people living in the household	Discrete		Socio-demographic	HH-level	X	X
HH has handwashing facility with soap	Household has fixed handwashing facility and soap	Binary		WASH	HH-level	X	X
Water available at place for handwashing	Water available at the place for handwashing	Binary		WASH	HH-level	X	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
HH purifies drinking water	Household uses any of the following methods of water purification: boil, add bleach, strain through cloth, use water filter, solar disinfect, wash jerry can with soap, or let stand and settle	Binary		WASH	HH-level	X	X
HH uses open freshwater source for drinking water	Household uses open freshwater source (swamp or lake) to collect drinking water	Binary		WASH	HH-level	X	X
Tap/borehole is main HH drinking water source	Tap/borehole is the main household drinking water source	Binary		WASH	HH-level	X	X
Home latrine	Household has home latrine (flush or pour flush toilet, covered pit latrine with/without privacy, or composting toilet)	Binary		WASH	HH-level	X	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Faecal contamination at water site closest to HH	Contamination occurring at closest water site from the household, based on whether human stool was observed at the site	Binary		Environmental	Ind.-level	X	X
Distance to closest water site from HH (km)	Euclidian distance to the closest water site from household GPS location (in km)	Continuous		Environmental	HH-level	X	X
Distance to closest water site from village centre (km)	Euclidian distance from the household to the closest water site from village centre (in km)	basic		Environmental	Vill.-level	X	X
Distance to closest public latrine from HH (km)	Euclidian distance from the household to the closest public latrine (km)	Continuous		WASH	HH-level	X	X
Distance to closest public tap/borehole from HH (km)	Euclidian distance from the household to the closest public tap/borehole (km)	Continuous		WASH	HH-level	X	X
Type of water site closest to HH	Type of water site closest to the household	Categorical	river, marsh, beach, swamp, pond, other	Environmental	HH-level	X	X
Water site type in village	Type of water site present in the village	Categorical	none, landing site, beach	Environmental	Vill.-level	X	X

Variable	Definition	Type	Levels	Type of variable	Level of variable	Candidate variable for water contact	Candidate variable for infection
Number of water sites per village	Number of water sites per village	Discrete		Environmental	Vill.-level	X	X
Number of water sites within 1 km of HH	Number of sites within 1km of the household	Discrete		Environmental	HH-level	X	X
Closest water site to HH in village	Whether the closest water site to the household is within the village where the household resides	Binary		Environmental	HH-level	X	X
Village-level schistosomiasis prevalence (KK)	Village-level prevalence by Kato-Katz (moderate 10-49% prevalence, high=50% prevalence)	Categorical	10-49%, 50+%	Biomedical	Vill.-level	X	X

Notes: Abbreviations: HH = household. Ind.-level = individual-level. Vill.-level = village-level. KK = Kato–Katz stool microscopy. Prop. = proportion. WASH = Water, sanitation, and hygiene.

Table S3.2. Key participant characteristics

Variable	No water contact (n= 1528)	Water contact (n= 1339)	P-value
Age (Mean $\pm$ SD)	21.3 $\pm$ 19.2	29.0 $\pm$ 15.9	<0.01
Age ² (Mean $\pm$ SD)	821.3 $\pm$ 1356.0	1091.8 $\pm$ 1058.1	<0.01
Female (%)	798 (52.2)	775 (57.9)	<0.01
Education level attained (%)			<0.01
none	249 (16.3)	182 (13.6)	
primary	1179 (77.2)	1009 (75.4)	
secondary or higher	100 (6.5)	148 (11.1)	
Enrolled at school (%)	526 (34.4)	274 (20.5)	<0.01
Religion (%)			0.06
no religion/other	17 (1.1)	30 (2.2)	
christian	1268 (83.0)	1101 (82.2)	
muslim	243 (15.9)	208 (15.5)	
Number of water contact activities (median [IQR])	0.0 [0.0, 0.0]	1.0 [1.0, 2.0]	-
Water contact duration per week (hrs) (median [IQR])	0.0 [0.0, 0.0]	8.0 [3.5, 17.5]	-
Recreational water contact (%)	0 (0.0)	78 (5.8)	-
Domestic water contact (%)	0 (0.0)	969 (72.4)	-
Water contact frequency per week (median [IQR])	0.0 [0.0, 0.0]	6.0 [3.0, 11.0]	-
Occupational water contact (%)	0 (0.0)	472 (35.3)	-
Occupation (%)			<0.01
none	1113 (72.8)	604 (45.1)	
fishing	25 (1.6)	226 (16.9)	
fishmongering	20 (1.3)	88 (6.6)	
farming	225 (14.7)	285 (21.3)	
other	145 (9.5)	136 (10.2)	
Home owned, not rented (%)	1297 (84.9)	1129 (84.3)	0.71
HH has handwashing facility with soap (%)	173 (11.3)	110 (8.2)	0.01
Home latrine (%)	1326 (86.8)	1133 (84.6)	0.11
Tap/borehole is main HH drinking water source (%)	774 (50.7)	739 (55.2)	0.02
HH uses open freshwater source for drinking water (%)	277 (18.1)	278 (20.8)	0.08
HH purifies drinking water (%)	354 (23.2)	293 (21.9)	0.44
Water available at place for handwashing (%)	265 (17.3)	199 (14.9)	0.08
People per HH (Mean $\pm$ SD)	3.6 $\pm$ 1.3	3.6 $\pm$ 1.4	0.96
Home quality score (Mean $\pm$ SD)	5.6 $\pm$ 1.1	5.7 $\pm$ 1.1	0.19
Faecal contamination at water site closest to HH (%)	719 (47.1)	577 (43.1)	0.04
Distance to closest water site from HH (km) (Mean $\pm$ SD)	0.4 $\pm$ 0.4	0.3 $\pm$ 0.3	<0.01
Distance to closest water site from village centre (km) (Mean $\pm$ SD)	0.3 $\pm$ 0.3	0.3 $\pm$ 0.2	<0.01

Table S3.2. Key participant characteristics (*continued*)

Variable	No water contact (n= 1528)	Water contact (n= 1339)	P-value
Distance to closest public tap/borehole from HH (km) (Mean $\pm$ SD)	0.3 $\pm$ 0.3	0.3 $\pm$ 0.3	0.76
Distance to closest public latrine from HH (km) (Mean $\pm$ SD)	1.1 $\pm$ 1.4	1.0 $\pm$ 1.4	0.03
Type of water site closest to HH (%)			<0.01
river	188 (12.3)	93 (6.9)	
marsh	590 (38.6)	412 (30.8)	
beach	491 (32.1)	566 (42.3)	
swamp	161 (10.5)	149 (11.1)	
pond	86 (5.6)	113 (8.4)	
other	12 (0.8)	6 (0.4)	
Number of water sites per village (Mean $\pm$ SD)	3.7 $\pm$ 1.4	4.0 $\pm$ 1.5	<0.01
Water site type in village (%)			<0.01
none	850 (55.6)	704 (52.6)	
landing site	173 (11.3)	104 (7.8)	
beach	505 (33.0)	531 (39.7)	
Number of water sites within 1 km of HH (Mean $\pm$ SD)	1.2 $\pm$ 1.3	1.3 $\pm$ 1.3	<0.01
Closest water site to HH in village (%)	566 (37.0)	605 (45.2)	<0.01
Village-level schistosomiasis prevalence (KK) = 50+% (%)	594 (38.9)	518 (38.7)	0.95

Notes: For binary variables, number and proportions [N (%)] are shown and Chi-squared tests were used to test for group differences. For all normally distributed continuous variables, mean and standard deviation (mean $\pm$ sd) are shown and two sample *t*-tests were performed to test for group differences. For all non-normally distributed variables, median and interquartile range (median [IQR]) are shown and Kruskal–Wallis Rank Sum Tests were performed to test for group differences. Abbreviations: HH = household. Prop. = proportion.

Table S3.3. Water contact per activity

Variable	Overall	Female	Male	P-value
Getting drinking water (%)	497 (17.3)	348 (22.1)	149 (11.5)	<0.001
Washing clothes with soap (%)	481 (16.8)	368 (23.4)	113 (8.7)	<0.001
Fishing (%)	354 (12.3)	39 (2.5)	315 (24.3)	<0.001
Washing clothes without soap (%)	152 (5.3)	115 (7.3)	37 (2.9)	<0.001
Washing jerry cans or other household items (%)	144 (5.0)	108 (6.9)	36 (2.8)	<0.001
Fishmongering (%)	104 (3.6)	90 (5.7)	14 (1.1)	<0.001
Swimming or playing (%)	78 (2.7)	32 (2.0)	46 (3.6)	0.018
Bathing with soap (%)	49 (1.7)	29 (1.8)	20 (1.5)	0.640

Table S3.3. Water contact per activity (*continued*)

Variable	Overall	Female	Male	P-value
Bathing without soap (%)	38 (1.3)	19 (1.2)	19 (1.5)	0.658
Collecting papyrus (%)	16 (0.6)	8 (0.5)	8 (0.6)	0.888
Collecting shells (%)	15 (0.5)	11 (0.7)	4 (0.3)	0.238

Notes: Number and percentage of participants with water contact [n (%)] per activity. Chi-squared tests were used to test for group differences.

Table S3.4. Water contact frequency per activity (in trips per week)

Variable	Overall	Female	Male	P-value
Trips to water collecting shells (median [IQR])	5.0 [3.0, 7.0]	5.0 [2.5, 7.0]	7.0 [6.5, 7.0]	0.13
Trips to water swimming or playing (median [IQR])	5.0 [3.0, 7.0]	5.0 [3.0, 7.0]	6.0 [3.0, 7.0]	0.78
Trips to water collecting papyrus (median [IQR])	5.0 [3.0, 5.5]	4.5 [2.8, 5.0]	5.0 [3.0, 7.0]	0.33
Trips to water bathing with soap (median [IQR])	5.0 [2.0, 7.0]	6.0 [2.0, 7.0]	3.0 [2.0, 7.0]	0.12
Trips to water fishing (median [IQR])	4.0 [2.0, 7.0]	7.0 [3.0, 7.0]	4.0 [2.0, 7.0]	0.06
Trips to water fishmongering (median [IQR])	3.5 [1.0, 7.0]	4.0 [1.0, 7.0]	3.0 [1.2, 6.8]	0.62
Trips to water bathing without soap (median [IQR])	3.0 [2.0, 7.0]	3.0 [2.0, 7.0]	4.0 [2.5, 7.0]	0.37
Trips to water washing jerry cans or other household items (median [IQR])	3.0 [2.0, 7.0]	3.0 [2.0, 7.0]	3.0 [2.0, 7.0]	0.97
Trips to water washing clothes without soap (median [IQR])	3.0 [2.0, 6.0]	3.0 [2.0, 5.0]	4.0 [3.0, 10.0]	<0.01
Trips to water washing clothes with soap (median [IQR])	2.0 [2.0, 5.0]	2.0 [2.0, 5.0]	2.0 [1.0, 5.0]	1.00
Trips to water getting drinking water (median [IQR])	7.0 [5.0, 14.0]	7.0 [5.0, 14.0]	7.0 [5.0, 12.0]	0.84

Notes: Median and interquartile range (median [IQR]) are shown. Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

Table S3.5. Water contact duration per activity (in hrs per week)

Variable	Overall	Female	Male	P-value
Time spent fishing (hrs) (median [IQR])	4.0 [3.0, 4.0]	4.0 [2.5, 4.0]	4.0 [3.0, 4.0]	0.81

Table S3.5. Water contact duration per activity (in hrs per week) (*continued*)

Variable	Overall	Female	Male	P-value
Time spent collecting papyrus (hrs) (median [IQR])	3.0 [2.0, 3.0]	3.0 [3.0, 4.0]	2.0 [1.8, 3.0]	0.01
Time spent collecting shells (hrs) (median [IQR])	2.0 [2.0, 3.5]	3.0 [2.0, 4.0]	2.0 [2.0, 2.2]	0.33
Time spent fishmongering (hrs) (median [IQR])	2.0 [1.0, 3.0]	2.0 [1.0, 3.0]	2.0 [1.0, 3.0]	0.98
Time spent washing clothes without soap (hrs) (median [IQR])	2.0 [1.0, 3.0]	2.0 [1.0, 3.0]	2.0 [1.0, 2.0]	0.90
Time spent washing clothes with soap (hrs) (median [IQR])	2.0 [1.0, 2.0]	2.0 [1.0, 2.0]	2.0 [1.0, 3.0]	0.61
Time spent swimming or playing (hrs) (median [IQR])	1.0 [1.0, 2.0]	1.0 [0.5, 3.0]	1.0 [1.0, 2.0]	0.92
Time spent bathing without soap (hrs) (median [IQR])	1.0 [0.5, 2.0]	1.0 [0.8, 2.0]	1.0 [0.5, 2.0]	0.49
Time spent washing jerry cans or other household items (hrs) (median [IQR])	1.0 [0.5, 1.2]	1.0 [0.5, 1.2]	1.0 [1.0, 1.2]	0.53
Time spent getting drinking water (hrs) (median [IQR])	1.0 [0.5, 1.0]	1.0 [0.5, 1.0]	1.0 [0.5, 1.0]	0.54
Time spent bathing with soap (hrs) (median [IQR])	0.5 [0.5, 2.0]	0.5 [0.5, 2.0]	0.5 [0.5, 2.0]	0.90

Notes: Median and interquartile range (median [IQR]) are shown. Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

Table S3.6. Water contact at high-risk time per activity

Variable	Overall	Female	Male	P-value
Washing clothes with soap at high-risk time (%)	167 (34.7)	124 (33.7)	43 (38.1)	0.460
Getting drinking water at high-risk time (%)	92 (18.5)	69 (19.8)	23 (15.4)	0.304
Washing jerry cans or other household items at high-risk time (%)	54 (37.5)	46 (42.6)	8 (22.2)	0.047
Fishing at high-risk time (%)	41 (11.6)	7 (17.9)	34 (10.8)	0.188
Washing clothes without soap at high-risk time (%)	35 (23.0)	26 (22.6)	9 (24.3)	1.000
Fishmongering at high-risk time (%)	21 (20.2)	21 (23.3)	0 (0.0)	0.067
Bathing with soap at high-risk time (%)	16 (32.7)	7 (24.1)	9 (45.0)	0.222
Swimming or playing at high-risk time (%)	13 (16.7)	5 (15.6)	8 (17.4)	1.000
Bathing without soap at high-risk time (%)	12 (31.6)	5 (26.3)	7 (36.8)	0.727
Collecting shells at high-risk time (%)	5 (33.3)	3 (27.3)	2 (50.0)	0.560
Collecting papyrus at high-risk time (%)	4 (25.0)	3 (37.5)	1 (12.5)	0.569

Table S3.6. Water contact at high-risk time per activity (*continued*)

Variable	Overall	Female	Male	P-value
----------	---------	--------	------	---------

Notes: Number and percentage of participants with water contact [n (%)] at high-risk time per activity. High-risk time was defined as water contact during peak cercarial shedding hours (10 a.m.–3 p.m.). Chi-squared tests were used to test for group differences.

Table S3.7. Comparison of variable selection between LRTs and BVS

variable	Infection status (LRT)	Infection status (BAS)	Water contact (LRT)	Water contact (BAS)
Age	X	X	X	X
Age ²	X	X	X	X
Female	X	-	X	X
Education level attained	X	X	X	-
Enrolled at school	X	-	X	-
Occupational water contact	X	-	-	-
Number of water contact activities	X	-	-	-
Occupation	X	X	X	X
HH has handwashing facility with soap	X	-	-	-
Home latrine	X	-	-	-
Faecal contamination at water site closest to HH	-	-	X	X
Distance to closest water site from HH (km)	X	-	X	-
Distance to closest water site from village centre (km)	-	-	X	-
Distance to closest public latrine from HH (km)	X	-	-	X
Type of water site closest to HH	X	-	-	X
Water site type in village	-	X	-	-
Number of water sites per village	-	X	-	X
Closest water site to HH in village	-	-	X	X
Village-level schistosomiasis prevalence (KK)	-	X	-	X

Notes: Comparison of variable sets selected using likelihood ratio tests (LRTs) at $p < 0.05$ with the variable sets selected via Bayesian variable selection (BVS) with marginal inclusion probabilities $p \geq 0.5$ (i.e., variables included in the median probability model).

Supplementary Figures

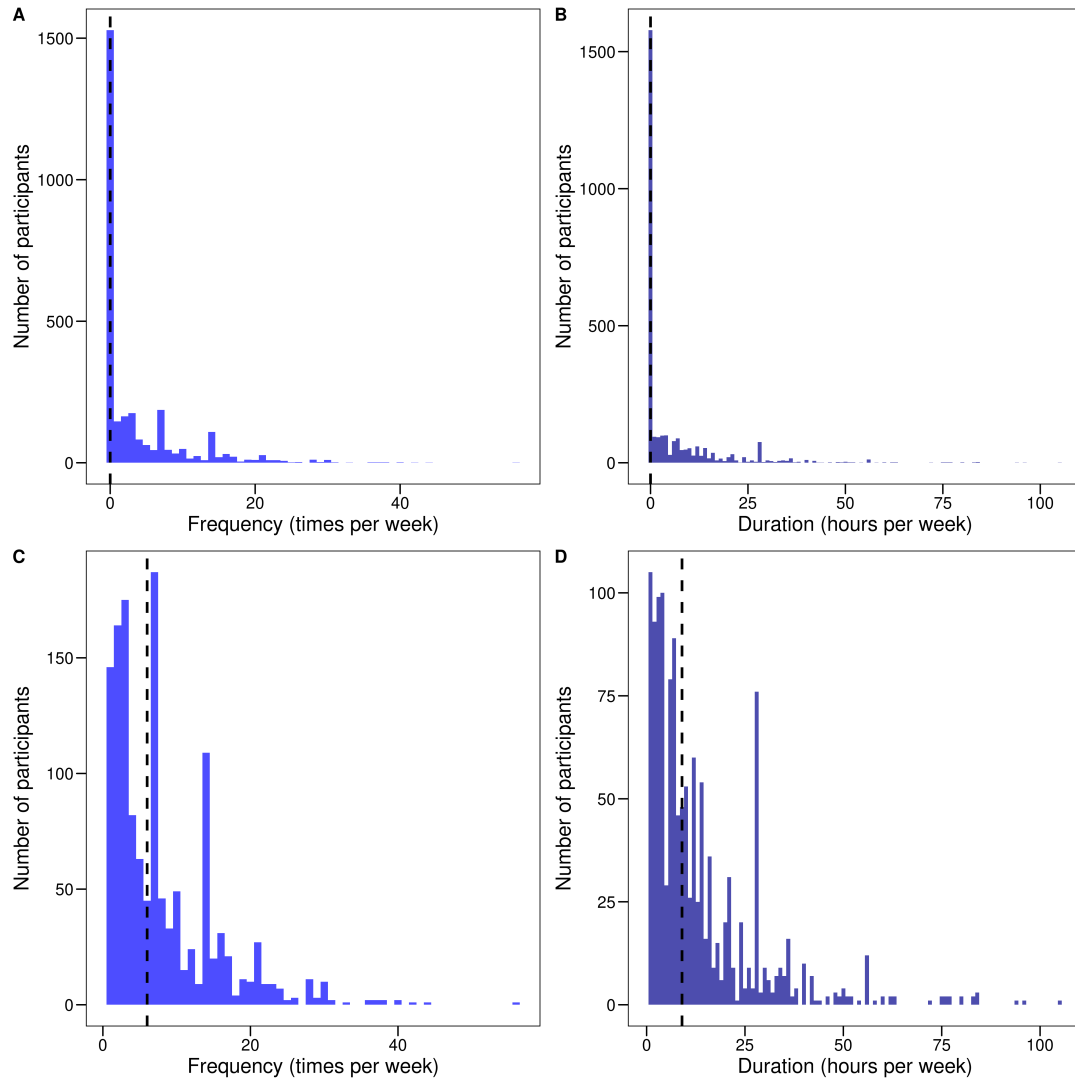


Figure S3.1. (A) Distribution of water contact frequency among all participants. (B) Distribution of water contact duration among all participants. (C) Distribution of water contact frequency among participants with water contact only. (D) Distribution of water contact duration among participants with water contact only. Dashed vertical line = median.

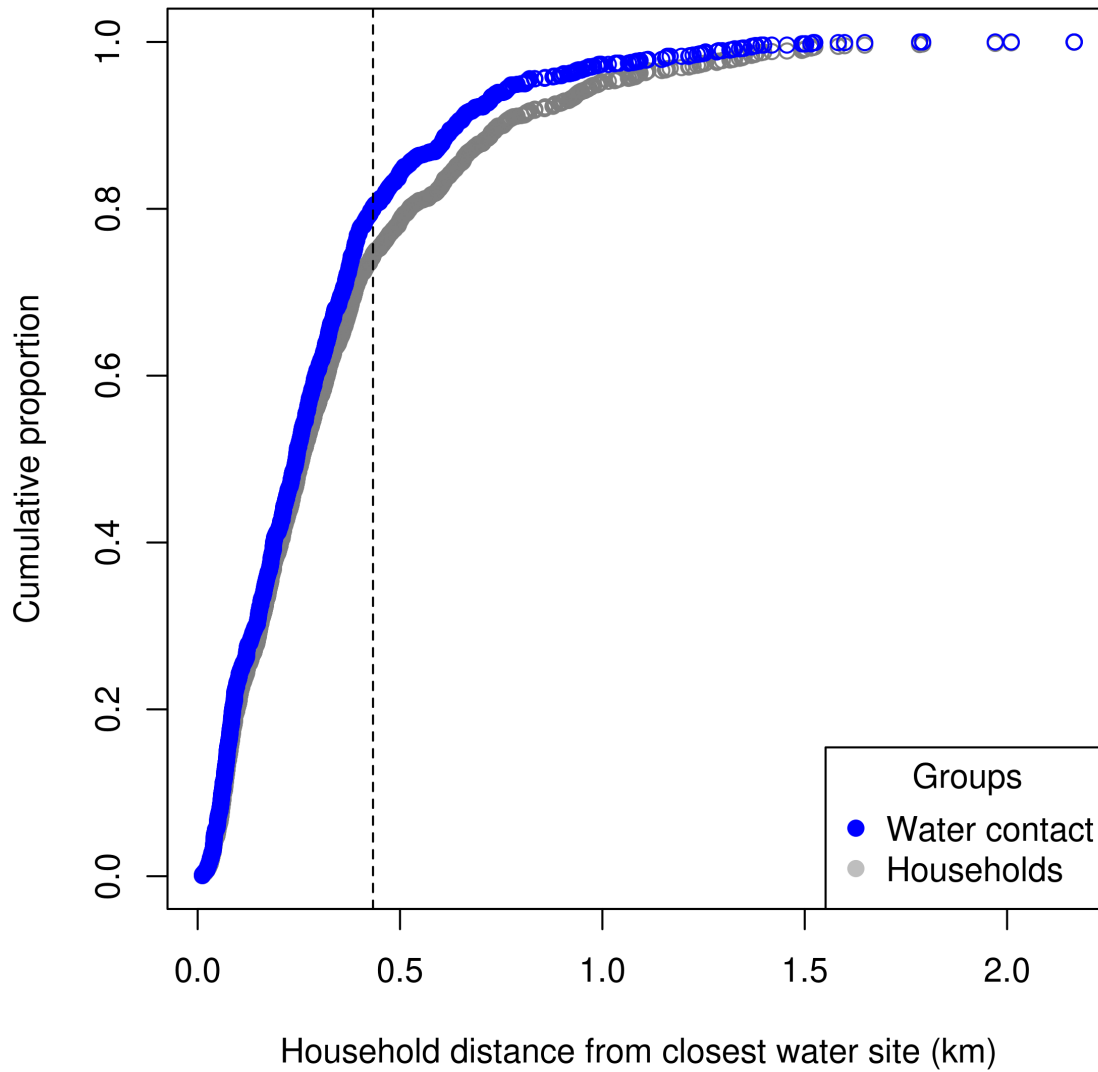


Figure S3.2. Cumulative proportion of all water contacts/participants by household distance to the closest water site ($n = 2867$). Dashed line shows the distance threshold where 80% of all individuals with water contact lived which is within 0.43 km from the shoreline. By comparison 74.3% of the overall study population lived within 0.43 km from the shoreline.

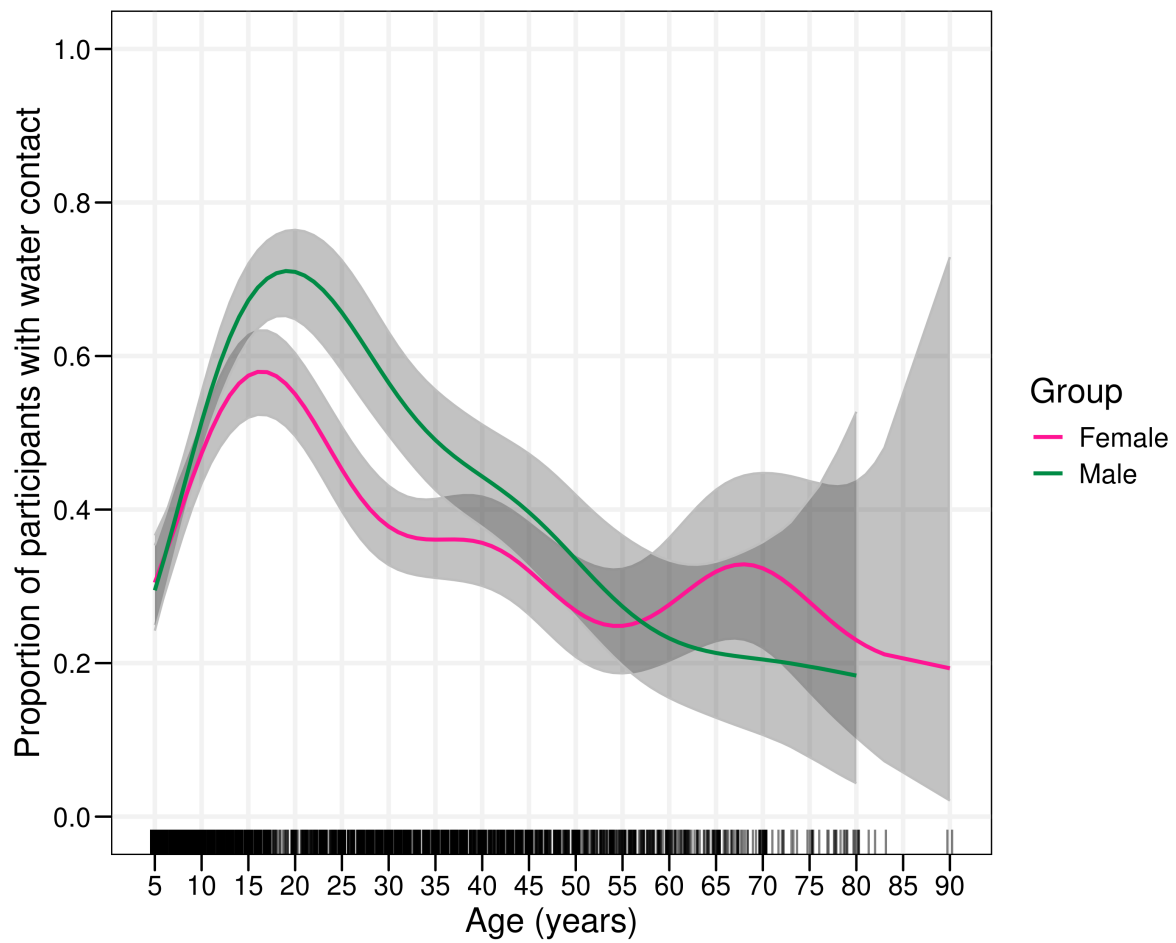


Figure S3.3. Generalised additive model predicting the proportion of participants with water contact over age and gender.

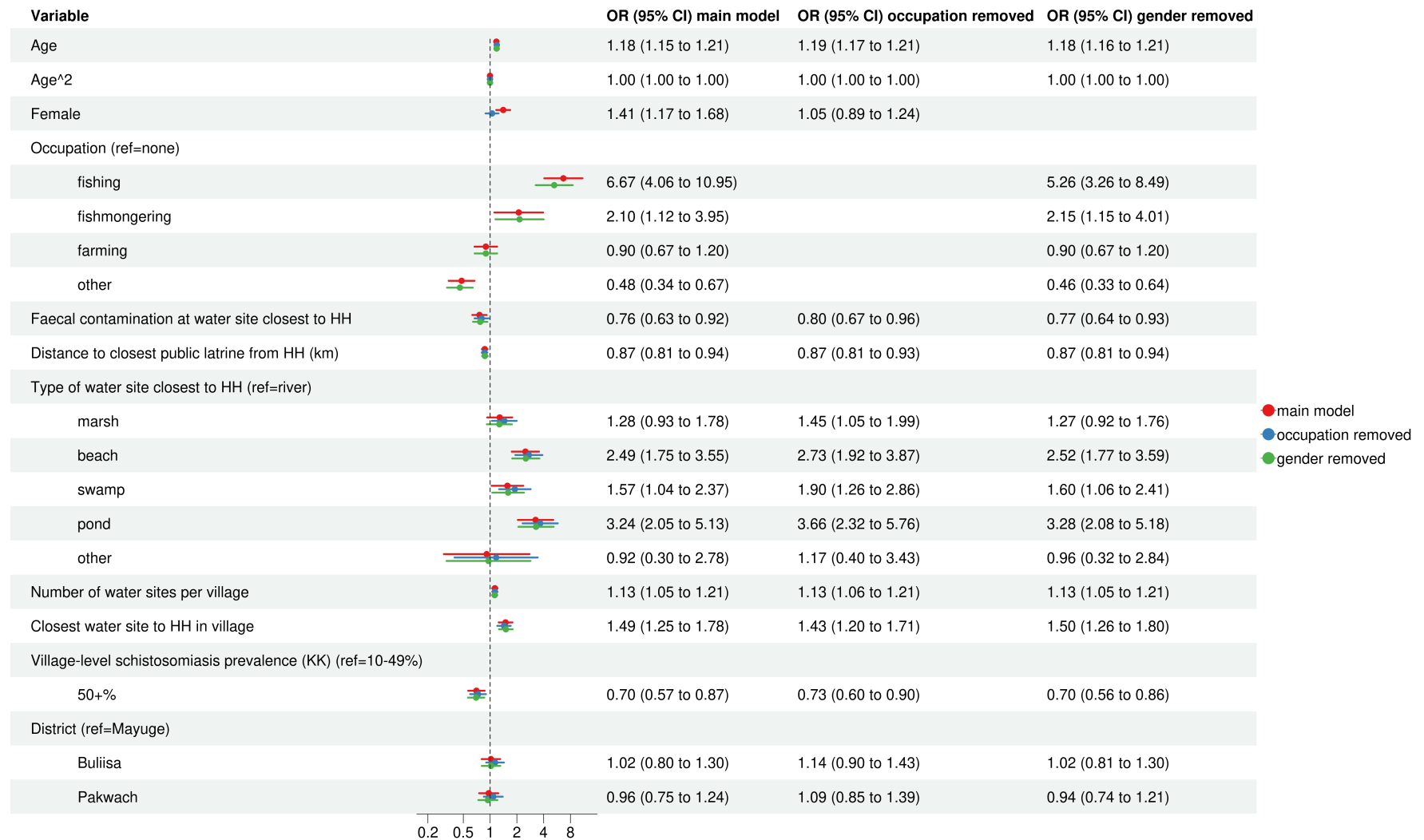


Figure S3.4. Logistic regression model predicting water contact (main model shown in Fig. 3.7) versus the same model with the occupation variable removed and gender removed, respectively ($n = 2867$ for all models). Standard errors clustered at the household level. Abbreviations: OR = odds ratio. CI = confidence interval.

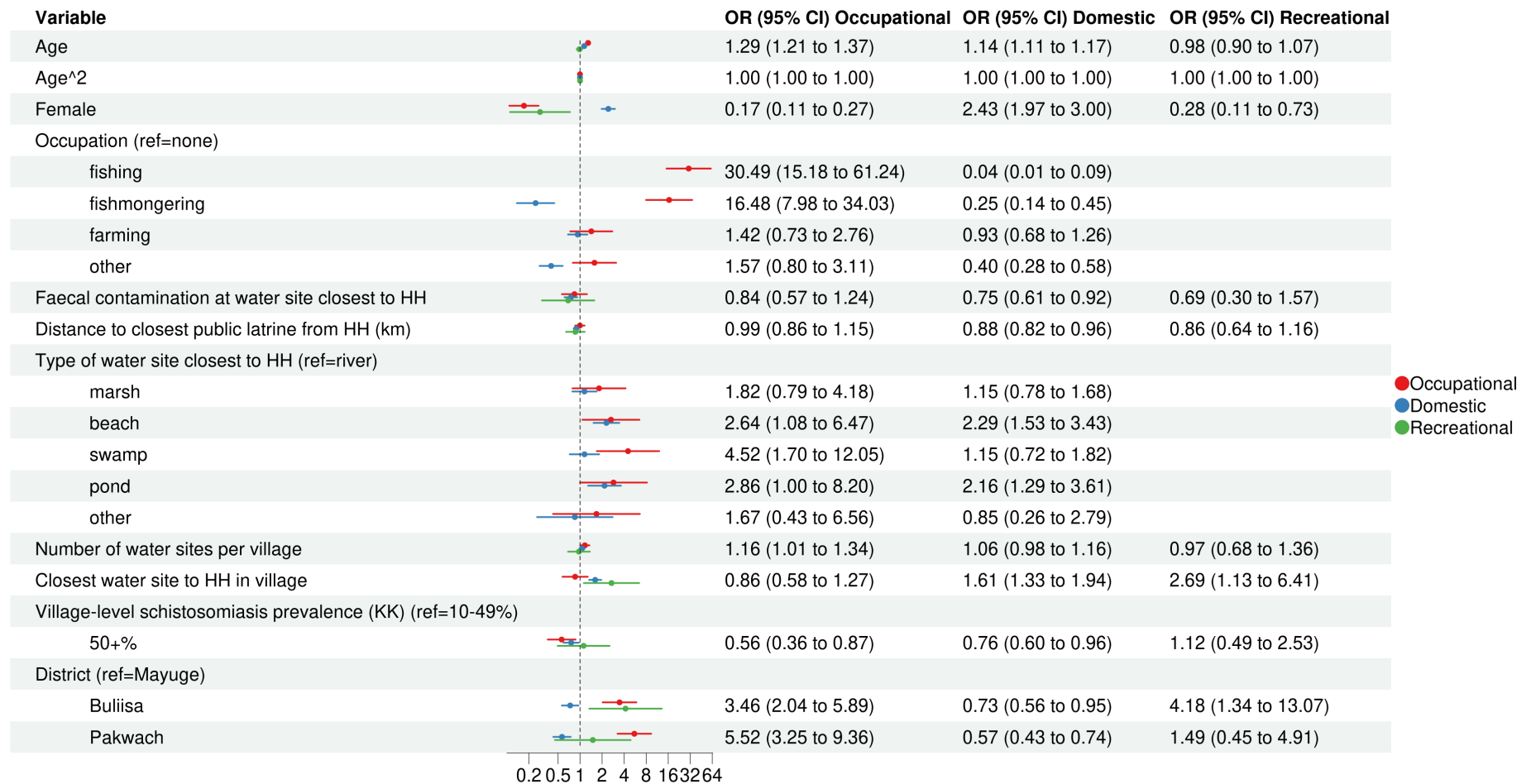


Figure S3.5. Results from separate logistic regression models with standard errors clustered at the household level predicting domestic, occupational, and recreational water contact among participants that do not engage in multiple types of water contact ($n = 2684$). As not all categories in the water site type and the occupation variables were sufficiently represented in the recreational water contact model, we removed these two variables when predicting recreational water contact. Abbreviations: OR = odds ratio. CI = confidence interval.

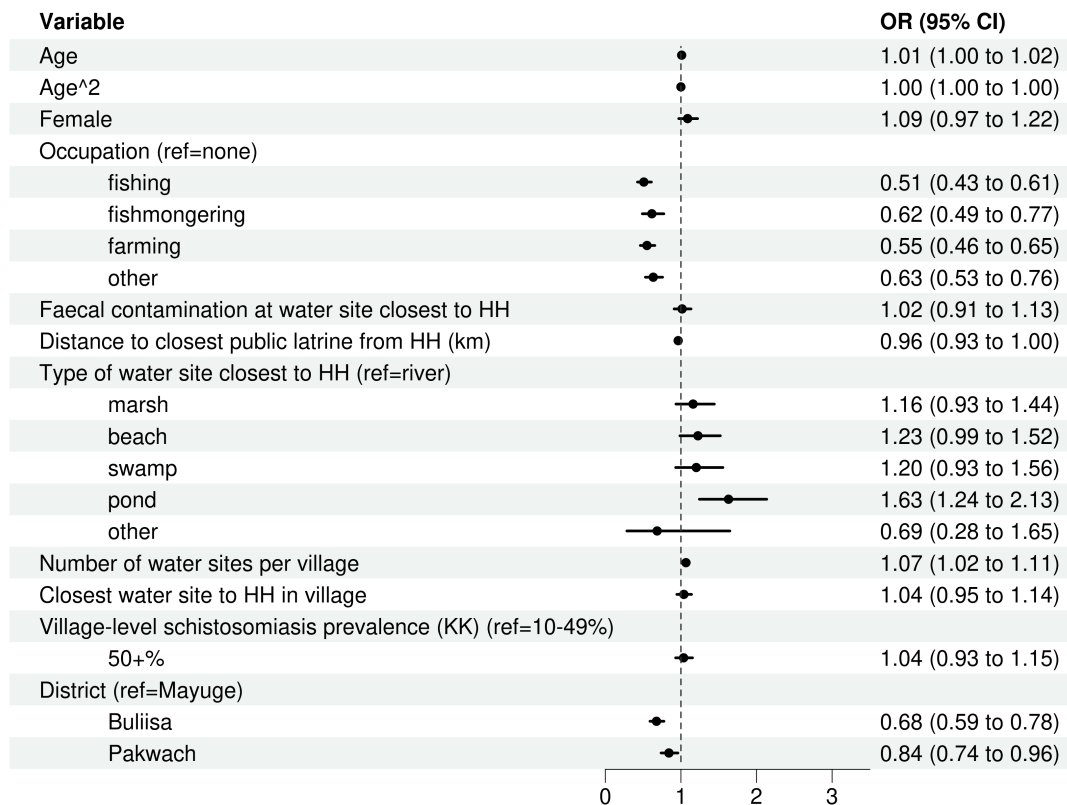


Figure S3.6. Negative binomial regression model predicting water contact frequency among all participants ($n = 2867$). Standard errors clustered at the household level. A zero inflation test suggested that the predicted number of zeros (1440) was not significantly different from the number of observed zeros (1528), ratio = 0.94, at the default tolerance of 0.1 of the `check_zeroInflation` function in the R package `performance`. Abbreviations: OR = odds ratio. CI = confidence interval.

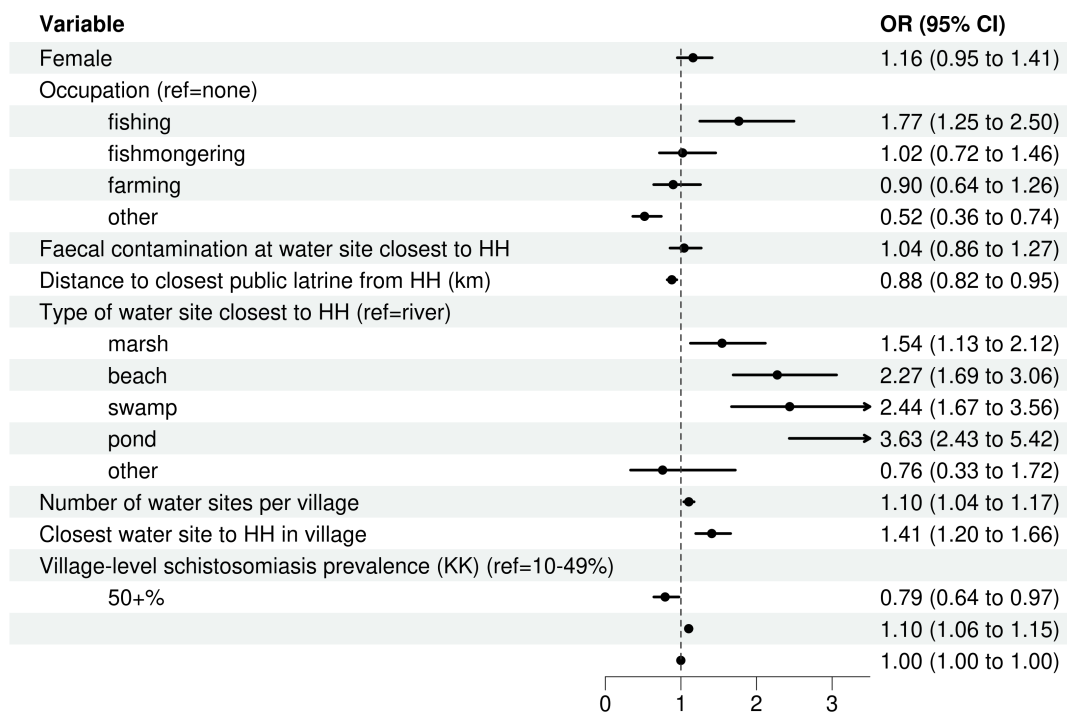


Figure S3.7. Negative binomial regression model predicting water contact duration among all participants ($n = 2867$). Standard errors clustered at the household level. A zero inflation test suggested that the predicted number of zeros (1479) was not significantly different from the number of observed zeros (1568), ratio = 0.94, at the default tolerance of 0.1 of the `check_zeroinflation` function in the R package `performance`. Abbreviations: OR = odds ratio. CI = confidence interval.

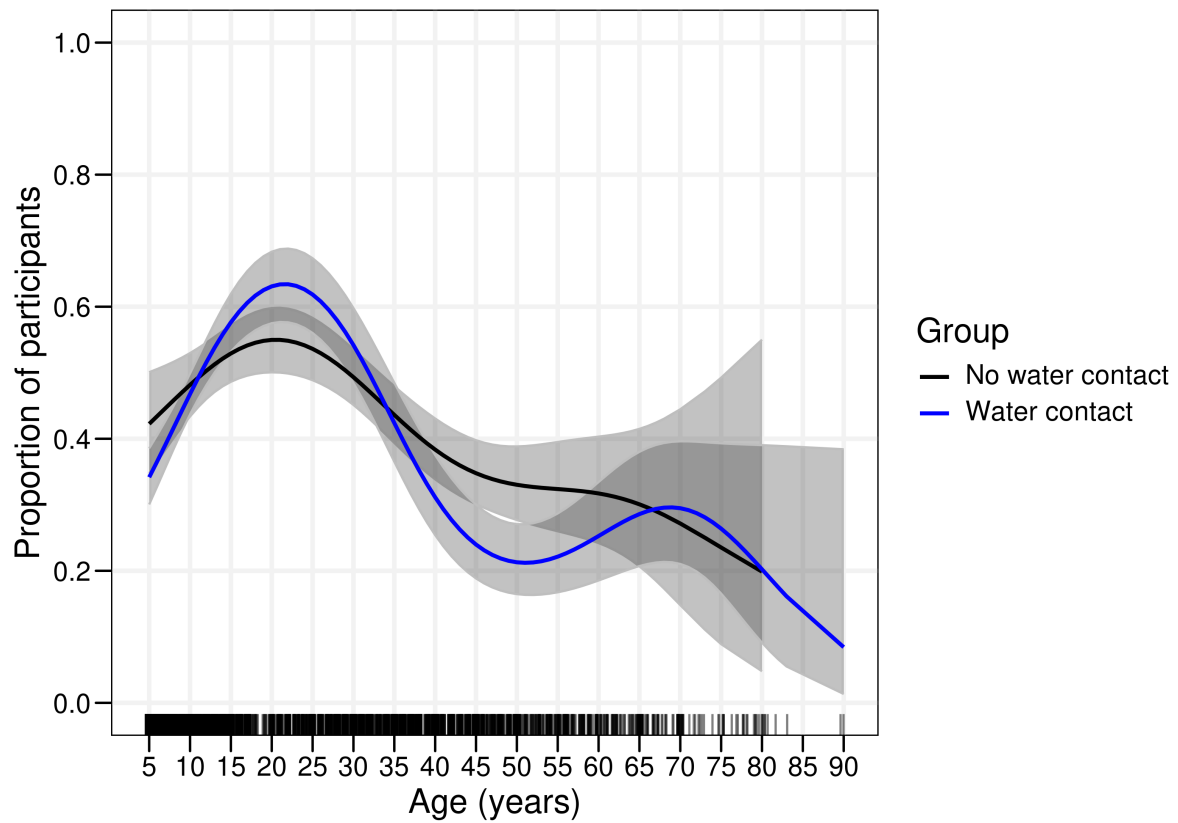


Figure S3.8. Generalised additive model predicting the proportion of participants with *S. mansoni* infection by water contact status and age (i.e., comparing infection outcomes in participants with current water contact vs participants without current water contact over age, $n = 2867$).

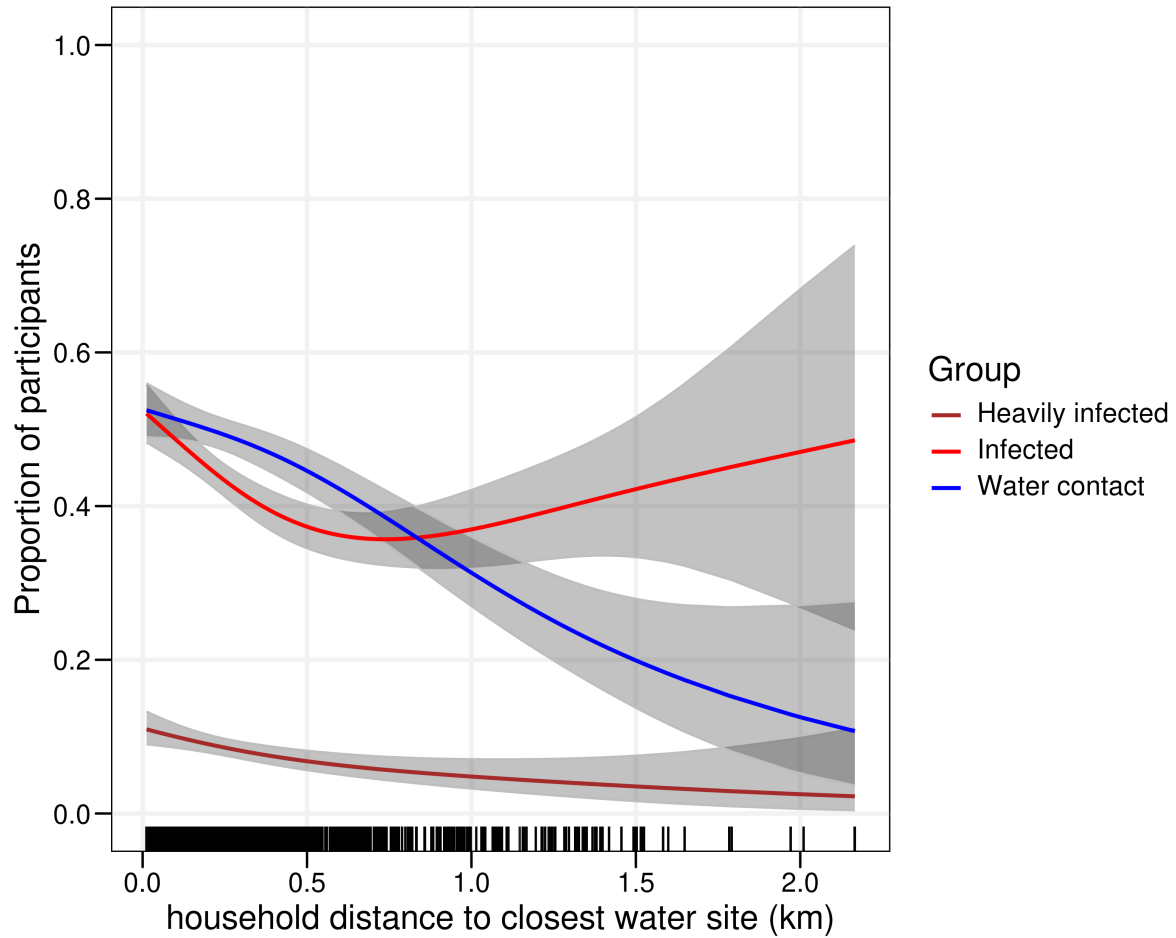


Figure S3.9. Generalised additive model predicting the proportion of participants with *S. mansoni* infection and heavy infection (≥ 400 eggs per gram of stool, by Kato–Katz microscopy) as well as the proportion of participants with water contact over household distance to the closest water site ($n = 2867$).

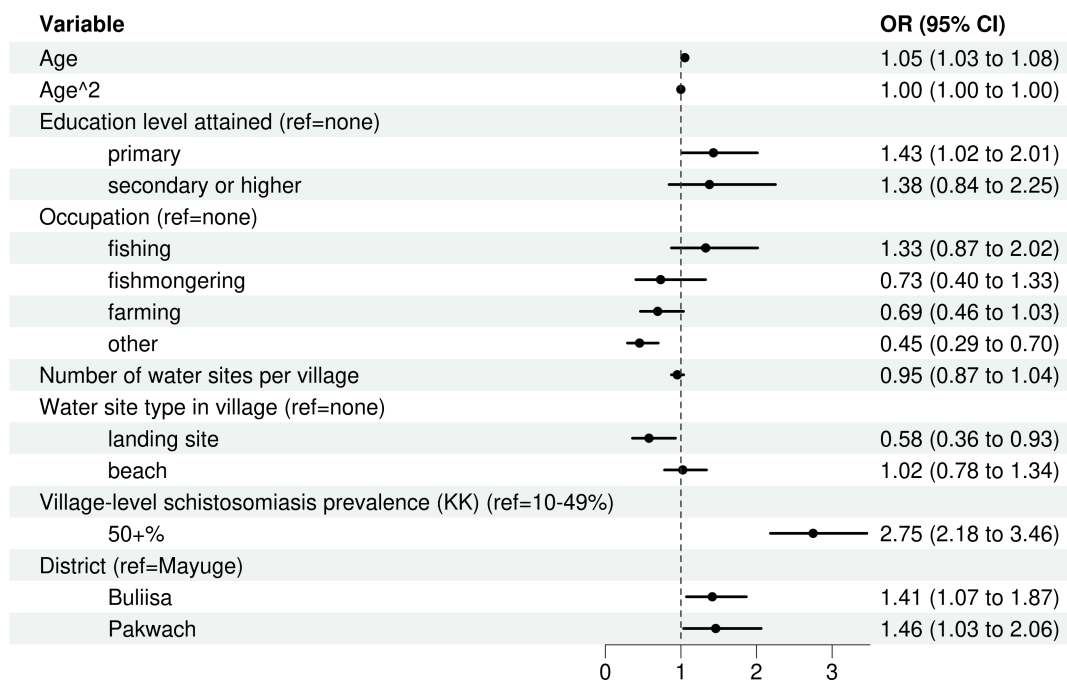


Figure S3.10. Zero-inflated negative binomial regression model predicting *S. mansoni* infection intensity among all participants ($n = 2867$). Intensity from Kato–Katz, measured in eggs per gram of stool. Standard errors clustered at the household level. A zero inflation test suggested that the predicted number of zeros (1531) was not significantly different from the number of observed zeros (1531), ratio = 0.94, at the default tolerance of 0.1 of the `check_zero_inflation` function in the R package `performance`. Abbreviations: OR = odds ratio. CI = confidence interval.

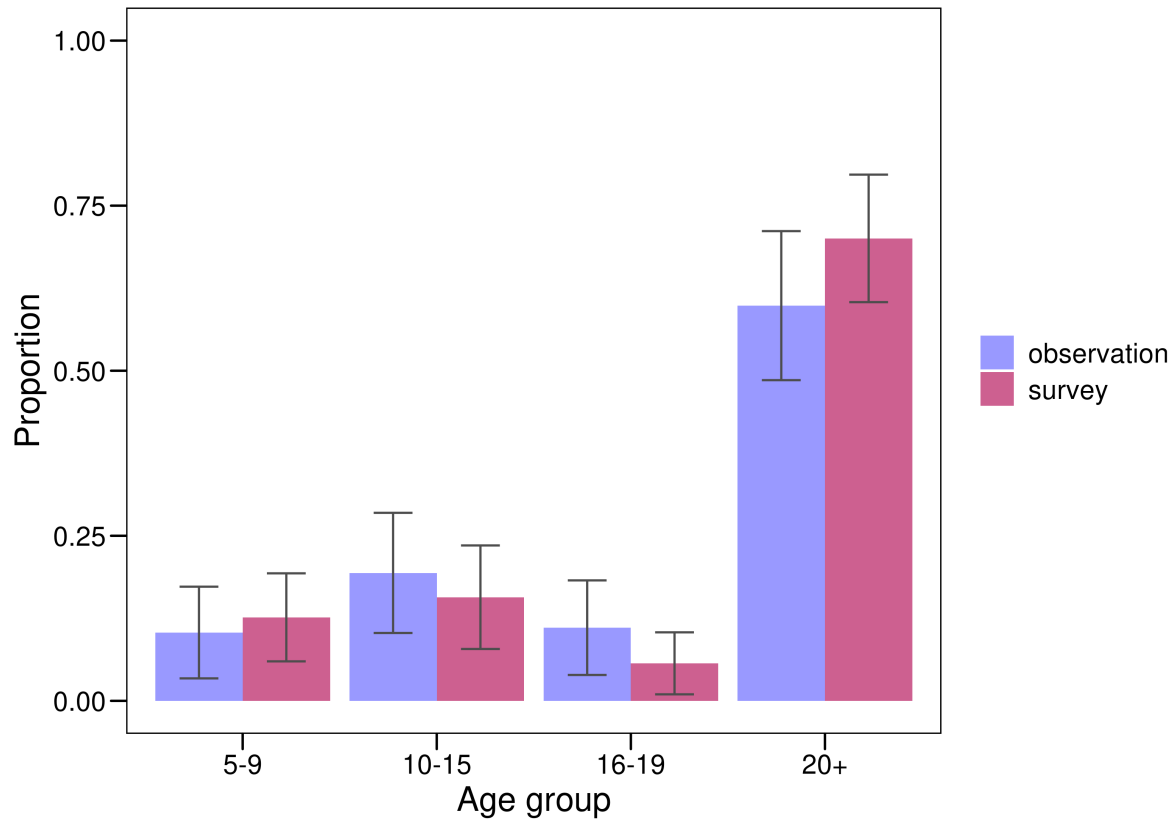


Figure S3.11. Comparison of age distribution in self-reported data with age distribution of direct observation data in 12 villages with both observation and survey data (survey data $n = 605$ and direct observation $n = 13,515$). Proportions represent the proportion of all water contacts which occur in each age group. Age groups were based on the categories used for MDA treatment and used because direct observation was conducted using these age groups. Whiskers represent 95% confidence intervals obtained using bootstrapping with 1000 repeats.

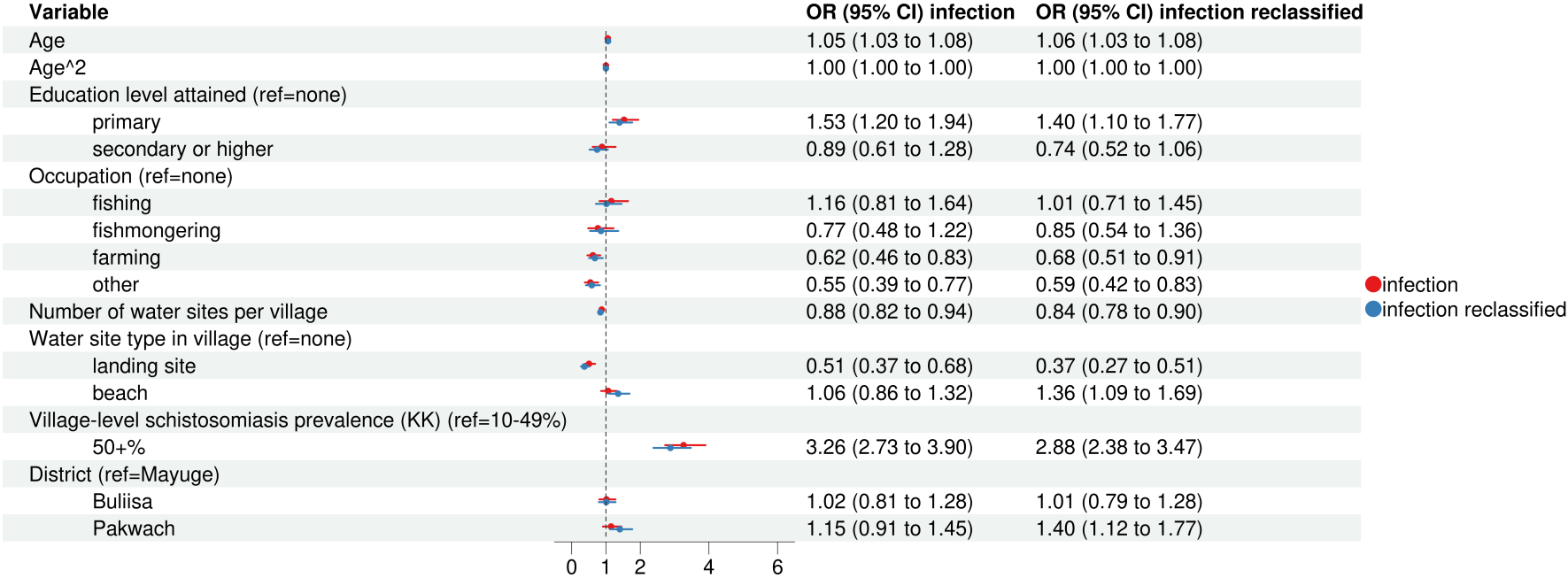


Figure S3.12. Separate logistic regression model predicting infection status based on Kato–Katz as well as reclassified Kato–Katz (where participants were recoded as infected whenever Kato–Katz infection status was negative but when the more sensitive POC-CCA diagnostic (positive band 1–3) indicated an infection, $n = 2867$). Standard errors clustered at the household level. Abbreviations: OR = odds ratio. CI = confidence interval.

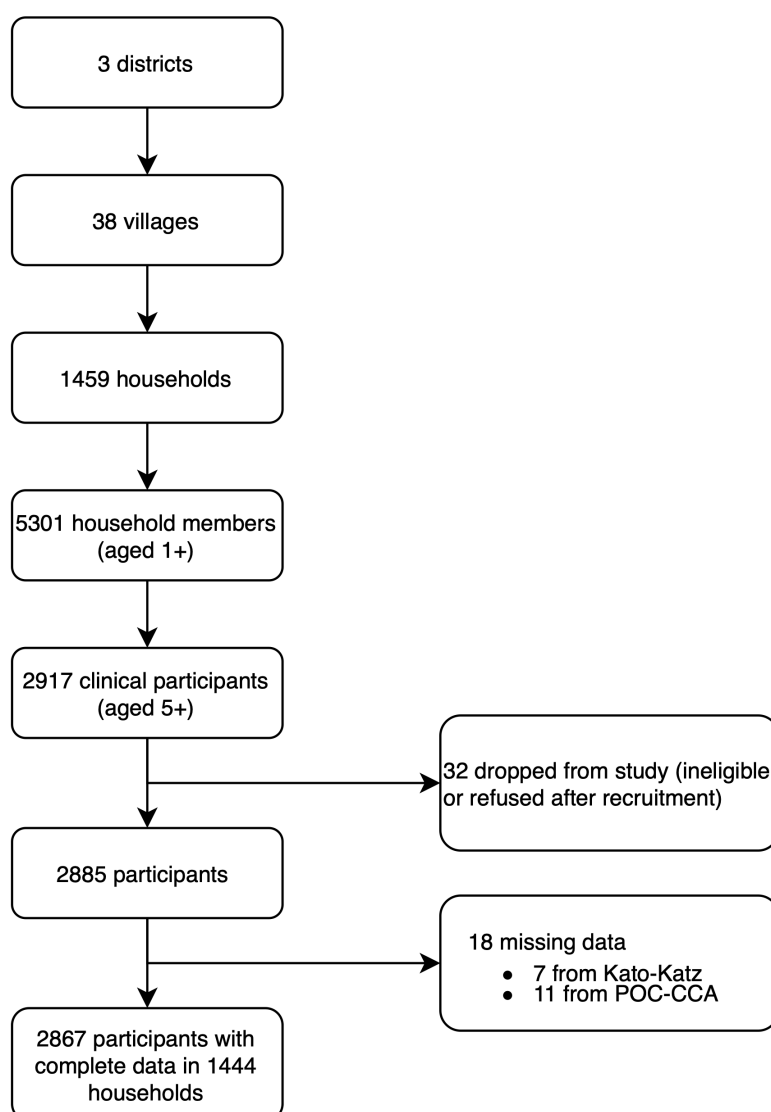


Figure S3.13. Participant flow diagram

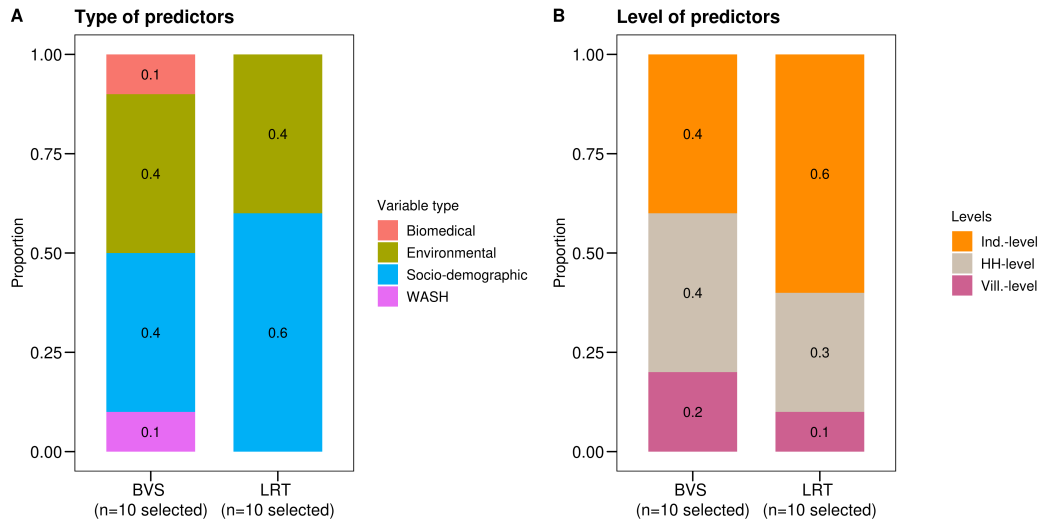


Figure S3.14. Comparison of variable selection results from BVS and LRTs for predicting water contact. (A) shows the percentage of selected variables from Bayesian variable selection (BVS, $n = 10$ selected variables) and likelihood ratio tests (LRTs, $n = 10$ selected variables) by type of variable. Variables selected using LRTs were solely socio-demographic and environmental variables, while BAS also selected water, sanitation, and hygiene (WASH) and biomedical variables. Among socio-demographic variables, age, age², gender, and occupation were consistently selected by BAS and LRTs. Current school enrolment status and educational attainment were only selected by LRT. Among environmental variables, contamination at the closest water site and presence of a water site within the village were consistently selected. Distance variables were only selected by LRTs (household distance to closest water site and village centre distance). BVS selected distance to the closest public latrine, the type of site closest to the household, contamination at the closest site, and village-level infection prevalence. No WASH variables were selected via LRT. In BVS, the WASH variables number of latrines per village and proportion of households using an improved drinking water source were selected. (B) compares selected variables from BVS and LRTs by level. In LRTs, 60% of selected variables were individual level, while the same figure was 40% for BAS. The proportion of village-level variables was twice as high in the BVS variable set (20%) compared to the set from LRTs (10%).

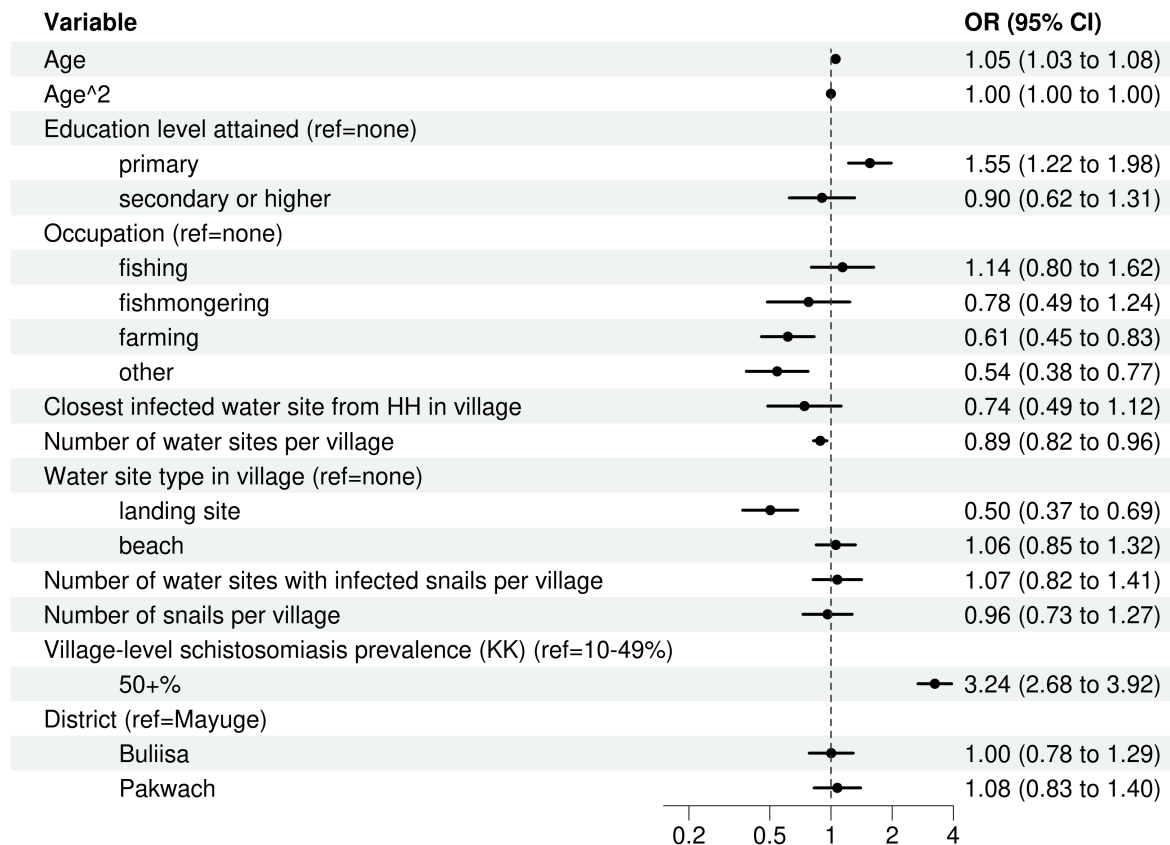


Figure S3.15. Logistic regression model predicting infection status among all participants using the main predictor variables plus a more extensive set of snail variables selected using Bayesian variable selection, see methods for details ($n = 2867$). Standard errors clustered at the household level. Additional snail variables selected via BVS were distance to the closest water site with infected snails and the number of infected snails at the closest water site. Abbreviations: OR = odds ratio. CI = confidence interval.

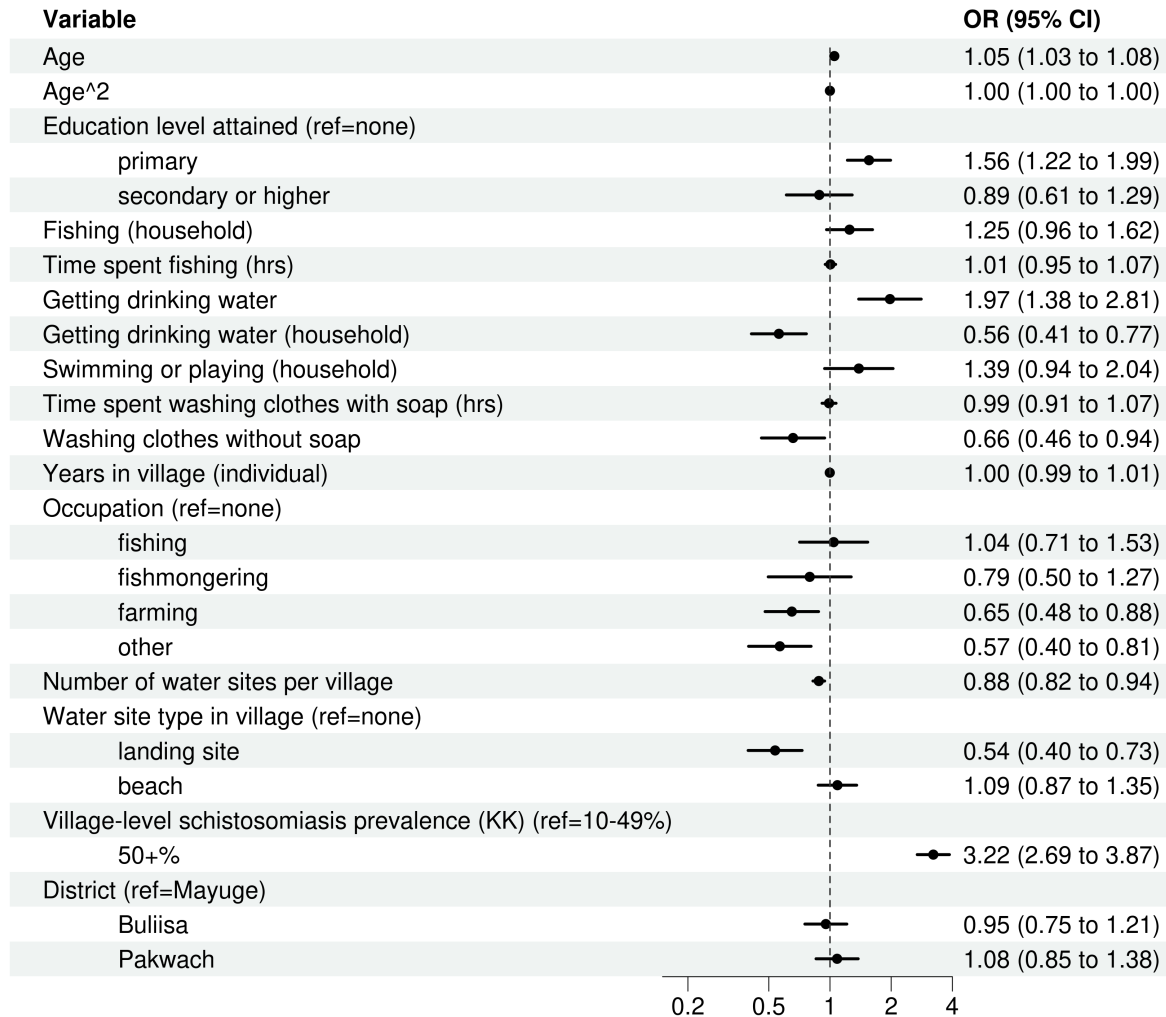


Figure S3.16. Logistic regression model predicting infection status among all participants using the main predictor variables plus a more extensive set of water contact variables selected using Bayesian variable selection, see methods for details ($n = 2867$). The six additional granular water contact variables selected via BVS were years of residence in the village, time spent on fishing, getting drinking water, household-level swimming or playing, washing clothes with soap, household-level fishing, time spent on washing clothes with soap, and household-level getting drinking water. Abbreviations: OR = odds ratio. CI = confidence interval.

Chapter 4

The Influence of Safe Water Supplies and Human Mobility on Water Contact Patterns in Rural Uganda: A Fine-Scale Spatial Modelling Study with Wearable GPS Loggers

4.1 Background

Globally, an estimated 703 million individuals did not have access to safe basic drinking water sources in 2022, failing to meet Sustainable Development Goal 6.1 of universal and equitable access to safe and affordable drinking water for all by 2030 [1]. 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 ponds [2]. The lack of access to safe water is associated with significant morbidity and mortality. In 2019, Prüssstün et al. estimated the global morbidity and mortality attributable to inadequate water, sanitation and hygiene (WASH) access across 183 countries. While the overall estimates were not specifically disaggregated by safe sanitation versus safe water-access attributable deaths, this analysis found that 637,000 deaths and 43.6 million disability-adjusted life years (DALYs) in sub-Saharan Africa may have occurred due to a lack of safe WASH [3]. Such modelling studies necessarily rely on quantitative data from field studies, but robust evidence to estimate the health impacts of safe water supplies is currently lacking for many infectious diseases [3, 4].

One such disease is schistosomiasis, the focus of this thesis. Providing basic safe public

water supplies, defined here as either boreholes, taps or standpipes, could reduce exposure, as schistosome transmission is reliant on contact with open freshwater sources [5]. The World Health Organization (WHO) has emphasised the important role of WASH interventions in schistosomiasis control programmes [6]. Yet, the 2022 WHO schistosomiasis control guideline rated the certainty of evidence for WASH reducing infection prevalence as low [6]. Analyses of the population-attributable fraction of schistosomiasis to inadequate WASH have arrived at different estimates. Prüstün et al. estimated this fraction to be 43% while an expert survey conducted by the WHO estimated it to be 82% [3, 7]. A recent analysis of the disease burden attributable to WASH did not include schistosomiasis because of the moderate WASH levels (basic WASH) which were used to estimate exposure-outcome relationships [4]. Empirical evidence of the effect of safe water on schistosome infection comes from two studies by Grimes et al. [8, 9]. An observational school-level analysis of safe water access found an ecological correlation between safer water collection practices and lower schistosome infection prevalence in Ethiopia [9]. A meta-analysis of 39 studies indicated that access to safe water was associated with lower odds of schistosome infection (OR = 0.53, 95% CI: 0.47–0.61). However, there was considerable statistical heterogeneity in the meta-analysis ($I^2 = 0.83$) and studies mostly reported unadjusted effect sizes which might be confounded by unobserved factors [8]. In contrast to observational evidence, a cluster randomised controlled trial in Zanzibar reported no additional benefit of behaviour change and WASH provision (urinals and safe washing platforms) when combined with biannual mass drug administration (MDA), compared to MDA alone [10]. As the effects of WASH on schistosome infection are mediated by changes in water contact behaviour as well as immunological, physiological or age-related factors [11–13], it is challenging to empirically estimate the effect of WASH, especially if the effect is small or when the levels of water contact required for infection are very low.

Estimating the impact of safe water supplies on water contact is important because the theory of change stipulates that safe water supplies cause reductions in open water contact and that this results in reductions in schistosome infection prevalence [5]. However, previous studies often focused on infection as the primary outcome of interest [8]. Therefore, if the expected outcome does not occur and no water contact data are collected, studies are not able to understand whether

the null result is due to (1) safe water supplies not reducing open water contact or (2) these reductions not translating into lower infection prevalence. We address this gap by understanding whether safe water supplies can reduce (i.e. crowd out) open water contact. If crowding out exists, it needs to be better understood how and under which conditions it occurs as the crowding out effect of water supplies could be linear where each increase in levels of community access to safe water supplies is associated with a proportional decrease in levels of open water contact in the community. Alternatively, crowding out may also be non-linear if a critical threshold of community coverage is required before reductions in community water contact occur [14]. There may also be another type of non-linearity where—despite high levels of community safe water access—open water contact is not further reduced. Empirical evidence on whether such crowding out occurs and its magnitude are currently lacking.

Capturing crowding out behaviour is challenging, especially when using self-reported data, which is the most common measure of water contact in the context of schistosomiasis, as shown in Chapter 2. Self-reported data have limitations in eliciting detailed water contact duration and frequency information, which makes it difficult to quantify crowding out effects on the intensity of water contact. There also is a risk of reporting bias due to the desirability of over-reporting WASH usage when WASH interventions and health education campaigns that often accompany such interventions are implemented in a community [15].

Data from wearable GPS loggers provide an alternative to self-reports and can yield detailed, objective measures of water contact duration or frequency as well as usage of safe drinking water sources. Two pilot studies in Uganda and Cameroon have demonstrated the feasibility of using GPS loggers to derive measures of water contact, although reliant on small samples ($n = 24$ and $n = 97$) over short time periods (24 hours and 72 hours) and with varying buffer areas (30m and 100m) to quantify water contact [16–18]. Existing GPS logger studies have not used environmental mapping of water sites to assign individuals to sites and were thus unable to study which specific water sites people use. As such information is also typically not available in studies using self-reported data, studies are usually required to make assumptions, for instance assigning individuals to all water sites in a village or the closest site [19–22]. These simple assignment rules might not capture complex human behaviour and social norms. Importantly,

when safe water supplies and water sites are mapped, the same GPS data could be used to study both water site usage and tap/borehole usage behaviours using one objective data source and consistent definitions which is highly beneficial for studying crowding out behaviour.

GPS data also contain human mobility information. The typical range of mobility may influence which water sites and taps/boreholes people visit. However, the extent of human mobility in rural settings and whether the distance travelled to visit water sites and taps/boreholes increases with the range of mobility is currently unknown. The importance of mobility for schistosomiasis has been shown theoretically in mathematical models [23–25], especially for longer-range transmission. Yet, whether mobility is an important determinant of local water contact and usage of safe water supplies remains unclear due to the lack of studies with detailed human mobility data at a sufficiently granular scale. For pathogens such as Ebola virus, SARS-CoV-2, influenza and measles, human mobility data have been incorporated into gravity models to predict the spatial dynamics of infectious disease transmission [26–28]. Gravity models typically describe population-level transmission patterns on larger geographic scales, such as between regions, cities, or countries. Mobile phone data has been used to incorporate mobility information into gravity models [26, 29]. However, in rural low-income settings where phone ownership and cell tower density are low, resulting in more coarse data, gravity models using mobile phone data have been shown to perform poorly [30]. In such settings, individual-level GPS data could enable the development of more granular models to predict the usage of points of interest, such as transmission sites.

In this study, individual-level spatial decay models were developed to predict fine-scale behavioural dynamics relevant to schistosomiasis. A GPS logger study with 452 individuals was nested within the SchistoTrack cohort, which includes participants from rural villages across three districts in Western and Eastern Uganda. The main research question was: Are spatial decay models able to predict fine-scale water site usage and crowding out via tap/borehole usage? The following sub-questions were answered: What are patterns of water site usage, tap/borehole usage, and human mobility in rural villages? What is the relevance of household distance to water sites and taps/boreholes for predicting usage and duration of usage? Which socio-demographic or behavioural factors mediate the relevance of distance? Is there any

evidence for crowding out water site usage by tap/borehole access?

4.2 Methods

4.2.1 Study Setting and Participant Sampling

The GPS logger study was conducted as part of the SchistoTrack baseline in January and February 2022 across Pakwach, Buliisa, and Mayuge districts in Uganda (described in detail in Chapter 3). Among the 38 villages included in the cohort, 12 villages (four per district) were selected for this study based on their levels of water contact and occupational fishing. All villages had a large landing site. None of the selected villages had piped water. Participants were chosen from the set of clinical participants that had been randomly sampled from village registers during the establishment of the SchistoTrack cohort 2022 baseline ($n = 2885$). Approximately 50 participants per village were recruited, with a target of one adult-child pair per household to ensure a balance of age groups and a diverse representation of water contact 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.

GFC, ME, and I conducted 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 (UGX 10,000, approximately USD 2.73) in addition to any incentive in the study upon returning their GPS loggers at the end of the observation period.

4.2.2 GPS Loggers and Participant Flow

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 two consecutive weeks. This time period was chosen to capture activities during the day. Loggers were turned off from 8 p.m. to 5 a.m. to save battery using scheduled control, as loggers were not recharged in the field by participants. They were set to motion activation such that loggers turned off when participants did not move—also to save battery. Button control on devices was disabled such that participants were unable to switch them off. Devices were placed in waterproof pouches and worn around the neck using a lanyard (Fig. S4.1). Participants were instructed to wear the devices during all daily activities, including during water contact, and to remove them only at night.

Two hundred and fifteen loggers were purchased for the study with the aim to hand out 200 loggers per district, i.e., 600 in total. Five hundred eighty-five GPS logger participants were recruited, slightly below the target of 600. This was due mainly to device failure and loggers not being returned by participants after hand out in Pakwach. After excluding participants with less than two complete days of GPS data—excluding the first day of recording for all participants to reduce Hawthorne effects [31]—we included a final sample of 452 participants in the study.

4.2.3 GPS Logger Fieldwork and Training During the DPhil

As part of my DPhil, I was heavily involved in the implementation of the GPS logger study between October 2021 and March 2022. Leading up to the fieldwork, I conducted various tests with the GPS loggers. I tested several GPS devices in Oxford and recorded data during hiking and cycling trips to assess their accuracy. With the help of FS, I charged all 212 devices and initialised them by taking them outside for several minutes. Together with ME, I researched suitable pouches for the loggers, and we identified a transparent, resealable pouch as suitable to protect against immersion in water via tests in a bucket of water. We also bought coloured lanyards to ensure that household members always received different coloured lanyards to reduce the mixing up of the loggers. I was involved in deciding the GPS logger device settings and developing the data management procedures. In the field, I assisted with developing the sampling procedure for the GPS logger participants and selecting the GPS logger villages. Together with ME, I prepared the devices for the field (charging them, setting them to scheduled control, putting them in pouches, and attaching lanyards). ME and I, with approval

from GFC, selected the villages for the GPS study and handed out the devices in the chosen villages. I also helped collect all devices from participants, recorded missing devices, attempted to retrieve devices from individuals who did not show up, and facilitated their reimbursement following device hand-in. After the hand-ins were completed, I helped with pulling data from the devices. For each GPS logger, we manually extracted the data with i-gotU GPS software using a bash script and saved the data as a CSV file on a field laptop before uploading the files to the BMRC cloud directly in the field. I also prepared the handout for the next round of data collection as we used the same devices across three districts. Following fieldwork, I created a merged master dataset and cleaned the dataset, removing non-study periods and erroneous GPS points, and calculated GPS accuracy using the calibration data we had collected. In August 2023, I completed a two-week summer school on GPS movement analysis, ANiMOVE, offered by the University of British Columbia in Kelowna, Canada, together with staff from the Max Planck Institute of Animal Behaviour in Konstanz, Germany. This school introduced me to data cleaning and analysis techniques for GPS movement data, including the recursion analysis, which I am using in this chapter.

4.2.4 GPS Logger Testing and Validation

During fieldwork, ten loggers were put next to each other on the ground for 10 minutes 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 8.6m. GPS precision was estimated using the `uere` function in the `ctmm` package in R [32]. Our estimate of 8.6m 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 less than 10m and that significant cover moderately increased location error by 2.2m [33].

4.2.5 GPS Logger Data Retrieval

For 94 participants, no GPS data could be retrieved as loggers suffered from water damage despite the use of protective pouches and indication from the manufacturer, Mobile Action, that the devices were fully waterproof (Fig. S4.2). Due to corrosion of the metal contacts, these

loggers could not be pulled using a computer, as they were not recognised by the i-gotU GPS logger software. We undertook multiple steps to retrieve data from the affected loggers. We cleaned the metal contacts with ethanol and attempted to retrieve the data in the field. We transported all loggers to Oxford, cleaned them again with ethanol, and attempted to retrieve the data once more. At that point, there were still 36 GPS loggers with missing data; we opened their cases at the University of Oxford Materials Science Department, where a research technician resoldered the GPS logger chips onto a functioning GPS logger. This enabled us to retrieve data from 24/36 GPS loggers.

4.2.6 Water Site and Tap/Borehole Data

We combined GPS data with ground-truth data on water site locations of 143 water sites and locations of 63 taps/boreholes, mapped as part of the SchistoTrack study. Trained malacologists, guided by either a village health team member or the village chairman, collected the GPS location of all water sites used by the local population. The GPS point was taken while standing in the water, and sites were limited to 15m in length. If sites were larger than 15m, they were recorded as multiple separate sites. Malacologists also recorded all human activities typically conducted at the site, based on information provided by the local guide. All public taps/boreholes within each study village were GPS-mapped by ME and I, also guided by a local, and initially with GFC. For this analysis, we only included working taps/boreholes (as described by village leaders and village health team members).

4.2.7 Deriving Water Contacts and Tap/Borehole Visits

A contact event with a water site and tap/borehole was defined as being within a 30m buffer of the GPS location of a water site or tap/borehole. We chose 30m to account for possible GPS position errors. GPS locations of sites and taps/boreholes were recorded via tablets (Lenovo Tab M8 (3rd Gen); Lenovo Group Limited, China) with 5–10m manufacturer-reported accuracy and GPS logger points with 8.6m accuracy, as reported above. All contact events with water sites and taps/boreholes were identified using recursion analysis, which is a well-established method in movement ecology [34, 35] and was implemented in the `recurse` package in R [36]. The

output from the recursion analysis was a list of all distinct water contact and tap/borehole usage events with an associated timestamp, number of revisits, contact duration, and water site and tap/borehole identifier. Compared to simply counting the number of points inside each buffer area, recursion analysis has the advantage that it reconstructs an approximated, linear trajectory and is therefore able to identify visitations with durations less than the GPS logger sampling frequency. This is particularly important given that water contacts are typically short [37].

As the SchistoTrack protocol did not allow sites to be larger than 15m, even when they were directly adjacent, we grouped such nearby locations into a single water site or tap/borehole for the analysis. The density-based spatial clustering of applications with noise (DBSCAN) algorithm was used and the optimal number of clusters was determined based on the gap statistic [38]. This collapsed the 143 water sites and the 63 taps/boreholes into 73 and 40 clusters, respectively, to which we then assigned the mean GPS cluster coordinates.

4.2.8 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 water site and tap/borehole usage. There were two reasons for this. First, field visits to all public latrines by ME and myself revealed that most latrines were locked (usually a community member held the keys), and they were not regularly used by the community. Second, among the 452 GPS logger participants, 343 (75.9%) lived in households that had a private latrine (i.e., a flush or pour-flush toilet, a covered pit latrine with/without privacy, or a composting toilet). Thus, most households could rely on private latrines which provide more privacy and convenience compared to public latrines [39, 40].

4.2.9 Dataset and Variables

Dataset. The dataset used to estimate spatial decay models were generated by calculating the Euclidean household distance from each household to all water sites and taps/boreholes (both used and unused) within the same district (Fig. 4.2).

Outcomes. For the models estimating any usage, the outcome was a binary indicator of whether

an individual used a water site or tap/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 water site or tap/borehole. The duration variable was generated by dividing the total duration of contact with each water site or tap/borehole by the number of distinct calendar days with GPS data for each individual.

Covariates. The main predictor variable was Euclidean distance from the household location to each water site or tap/borehole. We selected additional predictor variables based on their relevance to water contact patterns. The following individual-level variables were used: age, gender, occupation, self-reported water contact, activity match and mobility. Age was coded as a binary variable (< 15 years, ≥ 15 years). This cut-off was chosen because it was shown in Chapter 3 that water contact levels were low in individuals aged < 15 years and increased substantially thereafter. We also used gender (male/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 water contact [22]. Occupation of the household head (aged ≥ 18 years) was recorded. The occupation categories were fishing, fishmongering, farming, and other. For all participants aged < 18 occupation was coded as “none”. The self-reported water contact variable (which was the main outcome in Chapter 3) was also included. A binary activity match variable was generated for each individual–water site pairing. Whenever an individual conducted any of the 11 water contact activities from Chapter 3 and this same activity was reported to be done at a specific site, based on reports by the local guide, this was defined as an activity match. The water contact activities included getting drinking water, washing clothes with soap, washing clothes without soap, washing jerry cans or household items, washing clothes with soap, collecting papyrus, fishing, fishmongering, collecting shells, and swimming or playing. All individuals who reported getting drinking water were also assigned to having an activity match with taps/boreholes as this activity was amenable to being conducted at a tap/borehole. Apart from the activity match, water contact activities of an individual were not considered in this analysis as the GPS data does not contain any definitive information on the specific water contact activity an individual engaged in. We also used an individual’s mobility tercile as a covariate (see section below). At the household level, aside from distance

to the closest sites and taps/boreholes, a self-reported variable for whether the household used a safe drinking water source was also used. Taps/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 were different between study locations, a district-level categorical variable was included in some models. Households paid different prices for tap usage compared to boreholes (see results), therefore we also generated a variable indicating whether each public water supply was a tap or a borehole.

4.2.10 Human Mobility

Using the GPS logger data, we calculated the radius of gyration for each individual which serves as our measure for individual-level mobility. For each individual i , we calculate the radius of gyration $R_{g,i}$ around the household location using the following formula:

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

where $(\vec{r}_l - \vec{r}_h)^2$ represents the squared Euclidean distance from location l to the household location h . The radius of gyration provides a measure of mobility relative to an individual's household location and is widely used in human mobility modelling [41, 42].

4.2.11 General Spatial Decay Modelling Framework

The aim of this analysis was to model water site usage and tap/borehole contact behaviour of an individual based on the Euclidean distance of water sites and taps/boreholes from their household location. This approach was motivated by evidence that water contact and schistosome infection are waterbody distance-dependent [22, 43]. The results from Chapter 3 of this thesis, which used the full baseline data from the SchistoTrack cohort, and self-reported water contact data and closest site assignment, suggested that the decay in water contact over household distance was approximately linear. This finding is revisited here. It should be noted that the aim of this chapter was slightly different to the previous chapter where we predicted any water contact, assuming that individuals only use the closest water site. Here, the aim was to understand

whether site- and tap/borehole-specific usage patterns could be predicted without making the common assumption that individuals always use the closest site [20–22]. To estimate the usage probability of using a water site or a tap/borehole, we applied spatial decay models of the general form:

$$y = b_0 \cdot f(d)$$

where y is the probability of using a water site or tap/borehole and b_0 is a constant indicating the probability when the distance $f(d)$ is equal to 1. $f(d)$ represents the spatial decay function with d as the Euclidean distance between the household location and a water site or tap/borehole. For $f(d)$, two common decay functions were considered:

An **exponential decay** where:

$$f(d) = e^{-\lambda \cdot d}$$

and a **power-law decay** where:

$$f(d) = \frac{1}{(1 + \beta \cdot d)^{-\gamma}}.$$

In addition, we also considered a Hill function which is a more flexible function to model saturation and thresholds in the decay over distance:

$$f(d) = \frac{b_0}{1 + \left(\frac{d}{b_1}\right)^n}.$$

Identifying the decay function, $f(d)$, which best describes the observed data, allows for efficiently describing the decay process and predicting population-level water site and tap/borehole usage patterns.

The spatial decay models proposed here are closely related to gravity models. In a gravity model, pathogen transmission between two locations is determined by their respective population sizes and spatial distances whereby distance is modelled using a spatial decay function. Gravity models rely on aggregated mobility data to model population-level transmission patterns [27]. Here, we develop an individual-level model which, like the gravity model, can include both an attraction component and a spatial decay component. In contrast to the gravity

model where all locations are of one type (for instance cities), our model includes two different types of locations (households and water sites or taps/boreholes) and aims to predict how individuals (nested in households) use water sites or taps/boreholes at small spatial scales.

4.2.12 Spatial Decay Models of Water Site and Tap/Borehole Usage

We formulated several spatial decay models: models to predict any water site or tap/borehole usage (Eqs. 4.1–4.2), a model to estimate the effect of taps/boreholes on crowding out any water site usage (Eq. 4.3), models to predict the duration of water site or tap/borehole usage (Eq. 4.4), and a model to estimate the effect of taps/boreholes on crowding out water site usage duration (Eq. 4.5).

4.2.12.1 Any usage models

Global spatial decay model of water site and tap/borehole usage

$$P(\text{response}_{i,k} = 1) = b_0 \cdot f(d_{i,k}) \quad (4.1)$$

In this model, P is the probability that individual i uses a specific water site and tap/borehole k . P is a binary outcome modelled using a Bernoulli distribution with an identity link function to represent probabilities directly. As far as possible, we used non-informative priors. For instance, in the exponential decay model, we used uniform priors $\text{Unif}(0, 1)$ for b_0 and λ , the most conservative priors possible as b_0 and λ can only take values between 0–1. For all models predicting the probability of usage, we ensure that $P(\text{response}_{i,j,k} = 1)$ remains a valid probability (i.e., within the range $[0, 1]$). To this end, the product $b_0 \cdot f(d_{i,k})$ is constrained such that $0 \leq b_0 \cdot f_j(d_{i,j,k}) \leq 1$. This condition is enforced during model fitting.

Group-specific decay model of water site and tap/borehole usage

$$P(\text{response}_{i,j,k} = 1) = b_{0,j} \cdot f_j(d_{i,j,k}) \quad (4.2)$$

In contrast to the global model, both b_0 and $f(d)$ are allowed to vary by covariate j . The

covariates which were tested were described above in the “Dataset and Variables” section.

Crowding out model of water site usage

As water contact patterns may be influenced by the availability of taps/boreholes, we tested this hypothesis by developing a global spatial decay model with crowding out where the probability of water site usage was inversely related to the probability of tap/borehole usage:

$$P(\text{water site usage}_{i,j,k} = 1) = b_0 \cdot f_j(d_{i,j,k}) \cdot \alpha(1 - P(\text{tap/borehole usage}_i)). \quad (4.3)$$

Here, α quantifies the magnitude of crowding out, again using an uninformative prior $\text{Unif}(0, 1)$. The crowding out model was estimated in a two-step process where first, the tap/borehole usage probabilities were estimated for each individual and then used as a covariate in the crowding out model. We include the covariate j from the best-performing model water site usage model (see below for model selection procedure).

4.2.12.2 Usage duration models

Global water site and tap/borehole usage duration model decay model

Analogously to the global decay model of usage, the duration model predicts the mean water site and tap/borehole usage duration t using the following formula:

$$E(t_{i,k}) = b_0 \cdot f(d_{i,k}). \quad (4.4)$$

This model is a negative binomial model to account for overdispersion, where

$$E(t_{i,k}) \sim \text{NegBin}(\mu_{i,k}, \phi),$$

with $\mu_{i,k} = b_0 \cdot f(d_{i,k})$ as the mean duration, and ϕ as the dispersion parameter, capturing variability beyond what is expected under a Poisson model.

Extreme duration values can lead to exaggerated values of the dispersion parameter ϕ , making the model appear more overdispersed. Thus, all duration values $> 99.9^{\text{th}}$ percentile,

corresponding to a duration of > 123 minutes per day, were capped by setting these values to 123 minutes. For exponential duration decay models, we used an uninformative prior $\text{Unif}(0, 1)$ for λ and a normal prior $\text{Normal}(1, 9)$, truncated at zero and 30, for b_0 . This prior was chosen to match the mean and standard deviation of the data.

Crowding out model of water site usage duration

Similar to the crowding out model of water site usage, we built a duration model in which the duration of water site usage was crowded out by the probability of tap/borehole usage:

$$E(t_{i,j,k}) = b_{0,j} \cdot f_j(d_{i,j,k}) \cdot \alpha (1 - P(\text{tap/borehole usage}_i)) \quad (4.5)$$

where α represents the magnitude of crowding out.

Baseline models

In this analysis, logistic regression models were used as baseline models to benchmark the performance of the spatial decay models. These models followed the general form:

$$\log \left(\frac{P(\text{response}_{i,k} = 1)}{1 - P(\text{response}_{i,k} = 1)} \right) = \beta_0 + \beta_1 d_{i,k} + \beta_2 j_i \quad (4.6)$$

where β_0 is the intercept, $d_{i,k}$ represents the Euclidean distance to the water site or tap/borehole k , and j_i is another covariate. The baseline models were estimated as Bayesian logistic regression models with flat priors for all coefficients, meaning equal probabilities were assigned to all possible values of the coefficients.

4.2.13 Statistical Analysis Methods

Estimating Spatial Decay Models

We used Bayesian non-linear regression models implemented using the `brms` package [44] in R version 4.1.0 to predict the probability of an individual using a specific water site or tap/borehole. 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% credible intervals. All Bayesian models, except for those where entire districts were held out, were fitted on a training dataset containing 85% of the full data. In models where one district at a time was held out to evaluate performance on the other districts, the training data consisted of 150 individuals sampled with replacement from the full data to ensure the size of the training data was balanced across districts. Due to convergence issues in the Bayesian modelling, we restricted our data to all sites within 3000m of the households, which captured over 99.6% of all water sites and taps/boreholes used. As only seven individuals lived within < 20m distance of water sites or taps/boreholes, predicted values were reported for distances between 20–3000m. For all Bayesian models estimated via *brms*, a Hamiltonian Monte Carlo sampling algorithm via the Stan backend, employing the No-U-Turn Sampler (NUTS), was used [44, 45]. Four Markov chains with 2000 iterations per chain (1000 warm-up and 1000 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 visually assess chain mixing, and the `pp_check` function was used to assess agreement between predicted and observed response values.

Selecting the Best Fitting Models

To determine which decay function and covariates yielded the best model fit, Bayesian leave-one-out cross-validation was used, and the best model was identified based on the expected log pointwise predictive density (ELPD) for a new dataset—a Bayesian leave-one-out measure of out-of-sample predictive fit [46].

For the decay function, ELPD was compared between global Hill, exponential, and power-law decay models. To select the optimal covariate j for the decay model, one variable at a time (district, gender, age (< 15 years, $\geq$ 15 years), occupation category, improved household drinking water source, and mobility tercile) was used to build a more granular decay model. The ELPD among these competing models was compared to identify the best model.

To establish whether a global model with crowding out performed better than a model without crowding out, ELPDs between those models were compared.

Model Performance Evaluation

Model performance (predictive capacity) in absolute terms was evaluated conservatively by predicting individual-level site and tap/borehole usage behaviour, i.e., whether the model correctly predicted which specific sites and taps/boreholes each individual would visit. The area under the receiver operating curve (auROC) was used as the metric of model performance. Performance metrics were calculated using a separate test dataset (15% of the data and it was ensured that all districts were represented). To convert the response probabilities between 0 and 1 into a binary outcome that could be compared with the true outcome data, the optimal cut-point probability was determined using the `cutpointr` package, with the F1 score as the criterion [47]. The F1 score was chosen because it is suitable for imbalanced data. Here, 88.4% of water sites within 3000m were not used and 90.8% of taps/boreholes were not used. auROC of all baseline logistic regression models was reported and later compared to auROCs of the best-performing spatial decay models to provide information on performance differences in absolute terms.

Importantly, our performance evaluation is substantially more stringent compared to how gravity models are conventionally evaluated. Pearson correlation coefficients or R^2 between predicted and observed mobility flows are typically used to evaluate gravity models [48–50]. We also report R^2 here. However, we are not primarily interested in measures of aggregate performance but in the more difficult problem of whether individual-level usage patterns can be correctly predicted. This problem is more difficult because, for instance, if an individual visits a different water site than the one predicted, this introduces two classification mistakes from the perspective of individual-level performance evaluation while the more lenient criterion of whether this individual has water contact is still satisfied.

4.2.14 Out-of-Sample Prediction

The ability of spatial decay models to predict population-wide patterns of use of field-recorded water sites was demonstrated using the publicly available Global Google–Microsoft Open Buildings Dataset (V3), which has mapped over two billion structures globally [51]. Footprints of all buildings in the study villages in Pakwach were extracted. As there were few commercial or public buildings in the area, we assume that each building represents a household.

To generate population-wide predictions of water contact, we calculated the centroid location of each building (Fig. 4.9B) and input the centroid GPS coordinates into a global spatial decay model (Eq. 4.1), which had been trained on the GPS logger data. This produced out-of-sample predictions of water site usage for all households in the study villages in Pakwach.

4.2.15 Sensitivity and Robustness Analyses

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

As most water contact events are brief in duration (in terms of minutes), there is a risk of misclassification, as such water contact may represent a different type of activity, for instance passing by a water site without having actual 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 minutes). We then recalculated the spatial decay model predicting water site usage including only water contact events above these thresholds to understand how 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/distribution of loggers and participant non-compliance. Among those with GPS data, the number of GPS points per individual per day was used as an indicator to examine differences in wear non-compliance between demographic groups.

This study was conducted across three different districts, and behaviour patterns from one district may not be generalisable. To address this, data from two districts at a time were used to predict water site usage in the held-out district.

The number of individuals and observation days were down-sampled to determine the minimum number of individuals and observation days required to achieve good model performance.

4.3 Results

4.3.1 Study Overview

A total of 585 participants were selected from 12 out of 38 study villages, representing approximately 50 participants per village (Fig. 4.1). Among selected participants, one participant refused (reason unknown). Each of the 584 remaining study participants wore a wearable GPS logger for a period of two weeks, with location data recorded at two-minute intervals between 5 a.m.–8 p.m. local time.

We obtained what we targeted as the minimum required data (≥ 2 days of observation, excluding the first day) from 452 of the 584 sampled participants. Exclusions were due to missing data ($n = 132$), mostly due to device failure and a few loggers being lost by participants, as well as poor-quality data (< 2 days; $n = 10$). Among the 452 included participants, 183 (40.4%) were from Pakwach district, 164 (36.3%) from Buliisa district, and 105 (23.2%) from Mayuge district. In total, 1,281,455 GPS data points were recorded, with a median of 2912 data points per participant (interquartile range [IQR] 1543–3634). The median number of days with GPS data per participant was 10 (IQR: 10–13). As device failures occurred due to reuse of the same GPS devices, the median number of points was highest in Pakwach, the first district in this study (median = 3533, IQR: 3137–3835), followed by Buliisa (median = 2250, IQR: 1118–3548) and Mayuge (median = 1880, IQR: 942–2689). The same applied to the number of observation days which was highest in Pakwach (median = 13, IQR: 13–13), followed by Buliisa (median = 12, IQR: 7–14) and Mayuge (median = 11, IQR: 8–13).

Field-recorded GPS locations of water sites and taps/boreholes were used to derive water contact information from the logger data. Using 30-metre buffers around each water site and tap/borehole, all GPS traces passing through these zones were classified as water contact or tap/borehole usage episodes. A schematic of the data processing steps to generate the dataset used for modelling is shown in Fig. 4.2.

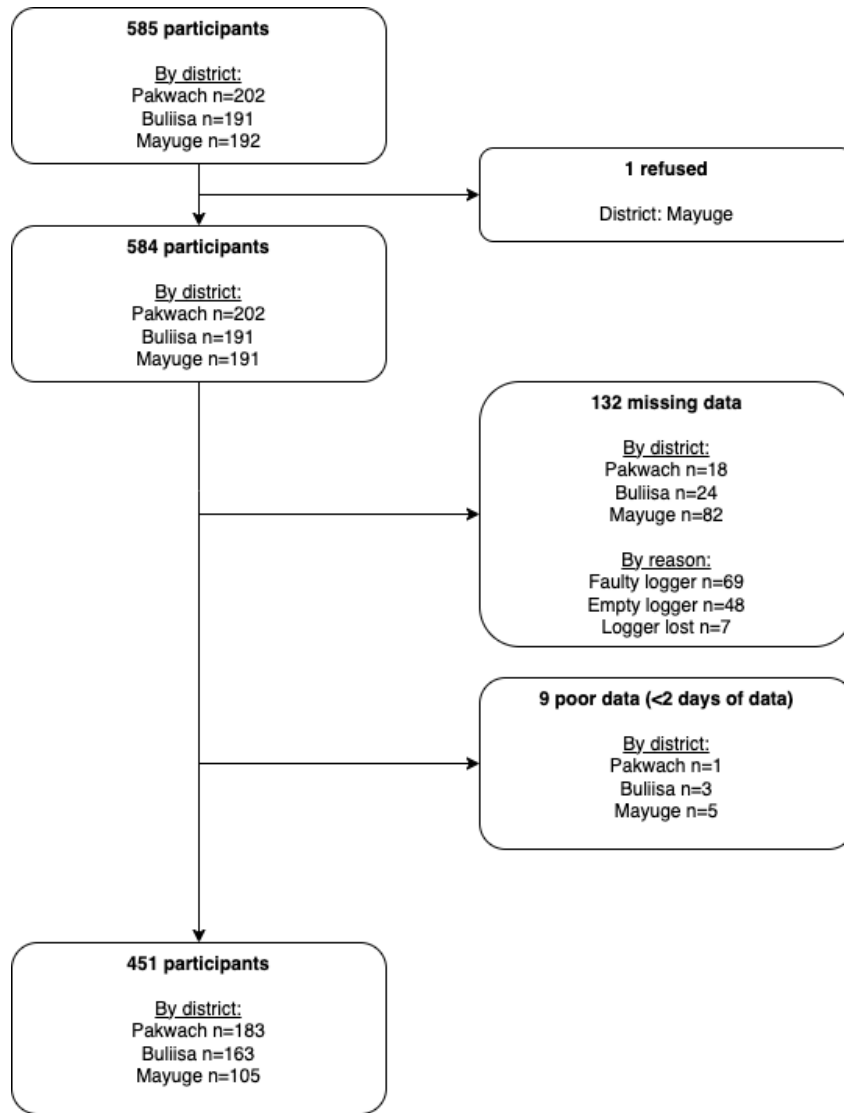


Figure 4.1. Participant flowchart. Loggers were classified as being faulty when no data could be retrieved due to complete device failure. Loggers were classified as being empty when loggers could still be pulled using a computer but contained no data.

4.3.2 Water Site Usage

Following the clustering of water sites, there were 15 water sites in Pakwach, 30 in Buliisa, and 28 in Mayuge. A total of 8249 episodes of water contact were observed during the study. Among the 452 participants, 64.4% (292/452) visited at least one water site. The median duration of water contact across all participants was 3 minutes (min) per day (IQR: 0–13.3 min). Among those who had water contact, the median daily duration was 9 min (IQR: 3–23 min), and the median number of different water sites visited was 1 (IQR: 1–2). Water contact patterns varied

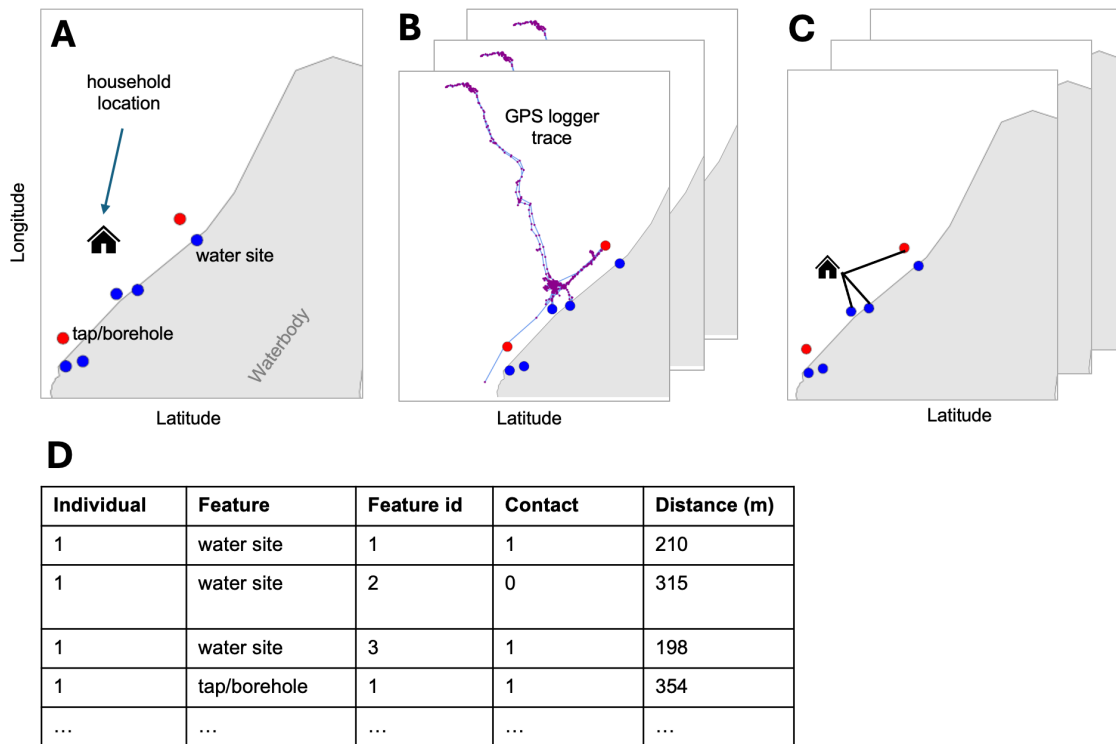


Figure 4.2. Schematic of deriving water site and tap/borehole usage from GPS logger data. (A) GPS locations of water sites and taps/boreholes. (B) Example GPS logger trace. (C) A network connecting an individual to all used sites. (D) A simplified data frame resulting from the processing.

across districts. In Pakwach, 53.6% of participants (98/183) used at least one water site. This number was 90.2% (148/164) in Buliisa and 46% (46/105) in Mayuge. Differences in diurnal patterns were also observed. In Pakwach, water contact peaked around midday, whereas in Buliisa and Mayuge, it was more evenly distributed throughout the day (Fig. S4.4).

4.3.3 Tap/Borehole Usage

There were 40 distinct locations with taps/boreholes. In Pakwach, there were two boreholes and five taps; in Buliisa there were three boreholes and 11 taps; in Mayuge there were two boreholes and 17 taps. The price for using these facilities was the same across districts; borehole usage cost UGX 1000 (USD ~ 0.27) per household per month and taps cost UGX 100 (USD ~ 0.03) per jerrycan. A total of 1181 tap/borehole contact episodes were observed, which was 85.7% fewer than the number of water contact episodes. Among all study participants, 33.6% (152/452)

visited at least one tap/borehole. The median duration of water contact across all participants was 0 min per day (IQR: 0–1 min). Among those who used taps/boreholes, the median daily duration was 1 min (IQR: 3–10 min), and the median number of different taps/boreholes visited was 1 (IQR: 1–1). Usage was highest in Buliisa, where 49.4% of participants used at least one tap/borehole (81/164), followed by Pakwach with 25.1% (46/183) and Mayuge with 21.9% (23/105). In Pakwach, tap/borehole usage was most common in the evening, whereas in Buliisa and Mayuge, it was more evenly distributed throughout the day (Fig. S4.4).

4.3.4 Combined Usage of Water Sites and Taps/Boreholes

Overall, 25.4% of participants (115/452) visited both water sites and taps/boreholes. Usage of water sites and taps/boreholes were weakly positively correlated (Spearman rank correlation [ρ_s] = 0.15, $p < 0.01$). There was a weak positive correlation between the duration of water site and tap/borehole usage, measured in min per person per day ($\rho_s = 0.17$, $p < 0.01$). Among participants with either water site or tap/borehole usage, there was a negative correlation between the duration of water site and tap/borehole usage ($\rho_s = -0.12$, $p = 0.02$, $n = 330$).

4.3.5 Demographic Differences in Water Site and Tap/Borehole Usage

Table 4.1 summarises the differences in water site and tap/borehole usage across demographic groups. Age and gender were not significantly associated with water site or tap/borehole usage. Occupation was linked to water site usage ($\chi^2 = 22$, $df = 4$, $p < 0.01$). In particular, fishermen and fishmongers were more likely to use water sites. By contrast, occupation was not relevant for tap/borehole usage.

Among people who used water sites, a higher proportion used an improved drinking water source (62.7%) than individuals who did not use any water sites (26.9%, $p < 0.01$). A higher proportion of those classified as using taps/boreholes based on the GPS data also self-reported use of improved drinking water sources compared to individuals who did not use taps/boreholes based on GPS data (67.3% versus 41.4%, $p < 0.01$).

GAMs suggested that water site and tap/borehole usage were dynamic across the life course. Duration of water site usage peaked around age 25 while tap/borehole usage duration decreased

with age (Fig. 4.3a). The duration of water site usage was higher than tap/borehole duration across all ages (Fig. 4.3a). Similarly, the average number of water sites used was higher than the number of taps/boreholes used across all ages (Fig. 4.3b).

4.3.6 Water Site and Tap/Borehole Usage Over Household Distance

The median Euclidean household distance to the closest water site was 225m (IQR: 82–412m), and the median distance to the closest tap/borehole was 251m (IQR: 93–664m). Distances were strongly correlated, with households closer to water sites tending to be closer to taps/boreholes as well ($\rho_s = 0.52$, $p < 0.01$). Among all districts, individuals in Buliisa lived closest to water sites and taps/boreholes (median water site distance = 69m [IQR: 47–89m]; median tap/borehole distance = 88m [IQR: 50–152m]). In Pakwach, individuals lived a median distance of 423m (IQR: 259–687m) from water sites and a median of 433m (IQR: 251–931m) from taps/boreholes. In Mayuge, the median household distance to the closest water site distance was 267m (IQR: 214–364m) and the median tap/borehole distance was 498m (IQR: 145–961m).

Overall, individuals travelled further to visit water sites compared to taps/boreholes; the median distance to a frequented water site was 232m (IQR: 106–425) whereas the median distance to a frequented tap/borehole was 112m (IQR: 50–270, Fig. 4.3c–d). For water contact, the closest site accounted for 45.7% (222/486) of all used sites and for 59.0% of total usage duration. The closest tap/borehole accounted for 87.2% (136/156) of all used taps/boreholes and for 88.6% of total usage duration. Visits to both water sites and taps/boreholes were highly concentrated in closest proximity to the household location (Fig. 4.3c–d) as 4.5% of total site visit duration and 4.7% of total tap/borehole visit duration was accounted for by sites and taps/boreholes located further than 1000m from the household.

4.3.7 Water Site and Tap/Borehole and Human Mobility

For each participant, the radius of gyration around the household location was calculated as a measure of human mobility. The median radius of gyration among participants was 380m (IQR: 151–1157m). Radii of gyration varied by district. Participants in Pakwach had the largest radius, with a median of 633m (IQR: 228–1454m), followed by those in Mayuge (median = 304m,

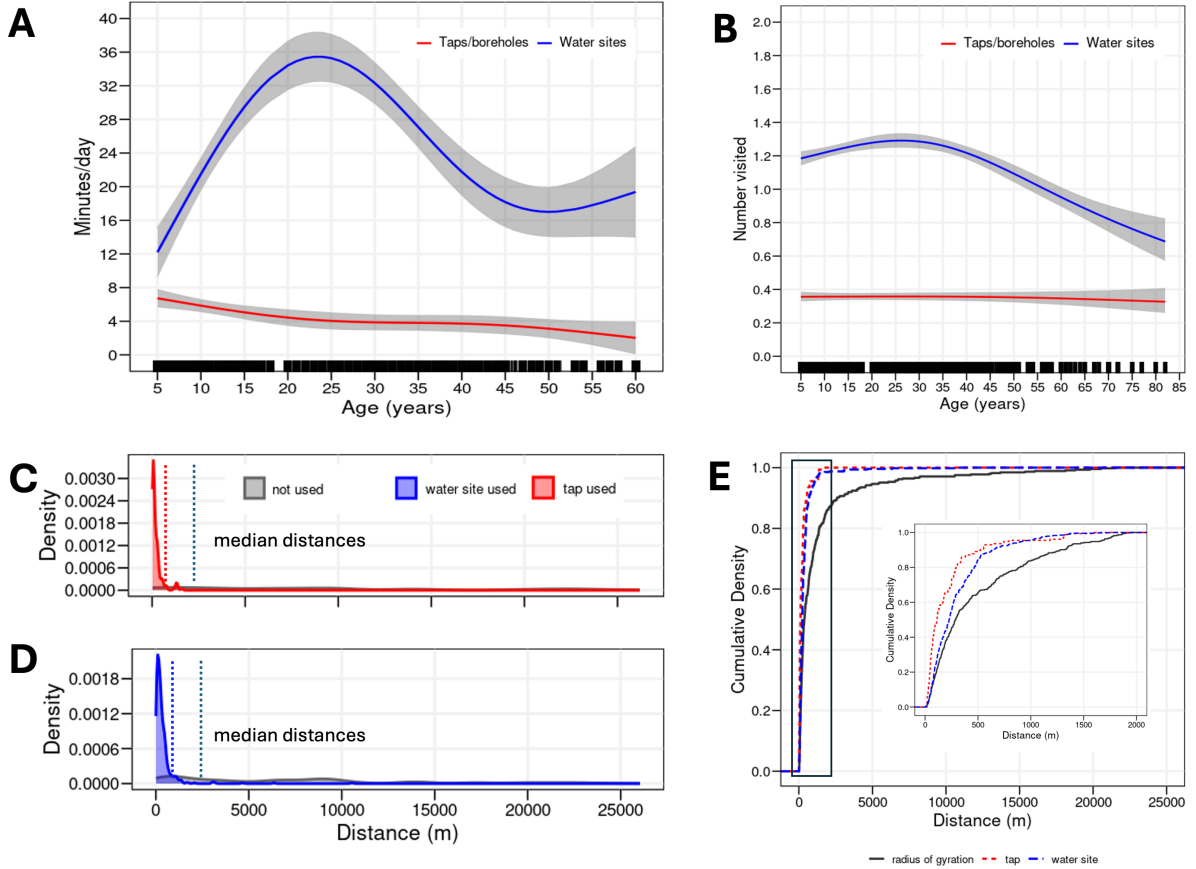


Figure 4.3. Water site and tap/borehole visits over age and distance. (A) Average duration of water site and tap/borehole usage over age in min per day, estimated using GAMs. (B) Average number of water sites and taps/boreholes used over age, estimated using GAMs. (C) Distribution of household distances for used versus unused taps/boreholes for all individuals. (D) Distribution of household distances for used versus unused water sites for all individuals. (E) Cumulative density of mobility (radius of gyration), tap/borehole usage, and water site usage over household distance.

Table 4.1. Summary of participant characteristics

Variable	Water site usage			Tap/borehole usage		
	No (n= 160)	Yes (n= 292)	P-value	No (n= 302)	Yes (n= 150)	P-value
Gender = male (%)	83 (51.9)	143 (49.0)	0.62	150 (49.7)	76 (50.7)	0.92
Age ≥ 15 years (%)	86 (53.8)	168 (57.5)	0.50	171 (56.6)	83 (55.3)	0.87
Occupation (%)			<0.01			0.60
none	95 (59.4)	164 (56.2)		176 (58.3)	83 (55.3)	
fishing	7 (4.4)	34 (11.6)		24 (7.9)	17 (11.3)	
fishmongering	2 (1.2)	20 (6.8)		15 (5.0)	7 (4.7)	
farming	46 (28.7)	47 (16.1)		65 (21.5)	28 (18.7)	
other	10 (6.2)	27 (9.2)		22 (7.3)	15 (10.0)	
Household drinking water source = improved (%)	43 (26.9)	183 (62.7)	<0.01	125 (41.4)	101 (67.3)	<0.01
Distance to the closest site (m) (median [IQR])	385.2 [248.3, 637.7]	91.6 [58.9, 249.6]	<0.01	312.8 ± 292.8	209.6 ± 200.6	<0.01
Distance to the closest tap/borehole (m) (median [IQR])	396.2 [135.2, 715.2]	179.4 [82.6, 404.9]	<0.01	515.7 ± 429.2	179.3 ± 270.7	<0.01
Radius of gyration (m) (median [IQR])	361.7 [170.9, 1180.5]	351.5 [128.6, 1071.5]	0.33	369.2 [152.6, 1092.6]	314.8 [134.7, 1065.3]	0.55

Notes: For binary variables, number and proportions [N (%)] are shown and Chi-squared tests were used to test for group differences. For all normally distributed continuous variables, mean and standard deviation (mean ± sd) are shown and two sample *t*-tests were performed to test for group differences. For all non-normally distributed variables, median and interquartile range (median [IQR]) are shown and Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

IQR: 152–698m) and Buliisa (median = 262m, IQR: 93–1222m). Radii were not significantly different between those using water sites and taps/boreholes compared to non-users (Table 4.1). For individuals using water sites, the median radius of gyration was 362 (IQR: 171–1181) versus 352 (IQR: 129–1072) for non-users ($p = 0.33$). For individuals using taps/boreholes, the median mobility range was 369 (IQR: 153–1093) versus 315 for non-users (IQR: 135–1065, $p = 0.55$). A moderate positive correlation was observed between the radius of gyration and the distances to used water sites ($\rho_s = 0.17$, $p < 0.01$) and used taps/boreholes ($\rho_s = 0.17$, $p < 0.01$). The cumulative distributions of mobility, distances to used water sites, and to taps/boreholes showed that visits to water sites and taps/boreholes were more focal than general human mobility (Fig. 4.3e).

The grouping of participants into district-specific terciles of mobility (bottom, middle, top; see Fig. S4.5) revealed socio-demographic differences (Table 4.2). Individuals in the top mobility tercile tended to be older; 69.3% were aged ≥ 15 years, whereas only 50.3% and 49.3% were aged ≥ 15 years in the middle and bottom mobility terciles, respectively ($\chi^2 = 16$, $df = 2$, $p < 0.01$). Fishermen and fishmongers had different levels of mobility. Among those individuals with high mobility 12.0% (18/151) were fishermen while only 6.6% (10/151) of low-mobility individuals were fishermen. For fishmongers the trend was reversed; 2.7% (4/150) of high-mobility individuals were fishmongers and 7.9% of low-mobility individuals (12/151) were fishmongers. While no relationship was found between gender and mobility overall, fishmongering and fishing occupations were highly gendered. Among those reporting fishing, 2.4% (1/41) were female, while among fishmongers 95.5% (21/22) were female.

Using GAMs, the average distance of water sites and taps/boreholes was shown to vary based on mobility. Individuals in the middle mobility tercile visited water sites approximately twice as far away (400m versus 200m) compared to individuals in the bottom tercile (Fig. S4.6a). Individuals in the top mobility tercile only visited sites significantly further away across ages 30–50, with a peak at 720m around age 42 (Fig. S4.6a). This age range coincided with the peak age of occupational fishing, as shown in Chapter 3. The average distance to frequented taps/boreholes linearly decreased over age for individuals in the bottom and middle terciles of mobility, while in the top tercile, distance was low across ages 10–15 and then similar to the

Table 4.2. Summary of participant characteristics

Variable	Mobility tercile (radius of gyration)			P-value
	Bottom (n= 151)	Middle (n= 150)	Top (n= 150)	
Gender = male (%)	74 (49.0)	77 (51.3)	75 (50.0)	0.92
Age $\geq$ 15 years (%)	74 (49.0)	76 (50.7)	104 (69.3)	<0.01
Occupation (%)				0.03
none	89 (58.9)	93 (62.0)	76 (50.7)	
fishing	10 (6.6)	13 (8.7)	18 (12.0)	
fishmongering	12 (7.9)	6 (4.0)	4 (2.7)	
farming	25 (16.6)	33 (22.0)	35 (23.3)	
other	15 (9.9)	5 (3.3)	17 (11.3)	
Household drinking water source = improved (%)	81 (53.6)	75 (50.0)	69 (46.0)	0.41
Distance to the nearest site (m) (Mean $\pm$ SD)	280.5 $\pm$ 260.6	311.4 $\pm$ 284.1	297.1 $\pm$ 298.6	0.62
Distance to the nearest tap/borehole (m) (Mean $\pm$ SD)	423.2 $\pm$ 425.0	398.4 $\pm$ 393.4	434.3 $\pm$ 434.3	0.74
Radius of gyration (m) (median [IQR])	96.8 [57.8, 159.9]	356.0 [251.9, 524.8]	1706.9 [1094.6, 3323.3]	<0.01

Notes: For binary variables, number and proportions [N (%)] are shown and Chi-squared tests were used to test for group differences. For all normally distributed continuous variables, mean and standard deviation (mean $\pm$ sd) are shown and two sample *t*-tests were performed to test for group differences. For all non-normally distributed variables, median and interquartile range (median [IQR]) are shown and Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

other terciles (Fig. S4.6b).

4.3.8 Logistic Regression Models for Predicting Water Site Usage

For every participant, the usage of all water sites within 3000m of their household location was predicted using variables conventionally applied to water contact prediction. The Bayesian logistic regression models presented in Table 4.3, with socio-demographic variables, behavioural variables, and distance-based predictor variables, serve as baseline models for further analyses.

Table 4.3. Performance of logistic regression models (baseline models)

Type of model	Predictors	Sensitivity	Specificity	AUC
Distance	Distance to water sites	0.594	0.929	0.854
Distance	Distance to water sites <430m (y/n)	0.764	0.745	0.755
Distance	Closest water site (y/n)	0.387	0.938	0.663
Behavioural proxies	Self-reported water contact (y/n), activity match (y/n), mobility tercile	0.821	0.324	0.590
Socio-demographic proxies	Gender, age (<15, ≥ 15 years), occupation category, improved household drinking water source (y/n)	0.726	0.416	0.544

Notes: Bayesian logistic regression models predicting GPS-derived water site usage. Models were trained on 85% of the data and performance evaluated on 15% of held-out data.

The worst-performing model used socio-demographic variables (gender, age, occupation category, and improved household drinking water source) to predict water site usage (auROC = 0.544). Another model used behavioural proxy variables for water contact (any self-reported water contact, whether self-reported water contact activities matched the activities typically performed at a water site, and the mobility tercile) to predict water site. This model (auROC = 0.590) outperformed the model with socio-demographic predictor variables with some sacrifice in sensitivity for improved specificity.

The three top-performing models relied on distance of the household to the water site to predict individual usage of water sites. A model in which all individuals were assigned using the closest water site to the household demonstrated an auROC of 0.663. Based on findings that over 80% of all water contact occurred in households within 430m of water sites (see Chapter 3), individuals were assigned to all water sites within 430m of their household location.

This model achieved an auROC of 0.755. The best-performing model was a logistic regression model using water site distance as a continuous variable to predict site usage (auROC = 0.854).

Table 4.4. Model selection for spatial decay models

Model	ELPD difference	Standard error	P-value
Water site models			
Mobility tercile	0.0	0.0	—
District	-7.2	11.2	0.52
Age	-20.5	6.4	0.001
Gender	-20.9	6.4	0.001
Activity match	-21.2	6.5	0.001
Occupation	-23.8	6.7	0
Tap/borehole models			
District	0.0	0.0	—
Activity match	-16.7	8.0	0.037
Mobility tercile	-36.0	11.0	0.001
Occupation category	-36.4	9.8	0
Gender	-37.9	10.4	0
Age	-38.0	10.5	0
Safe water supply type	-49.1	13.3	0

Notes: Model comparison using expected log pointwise predictive density (EPLD) to select the best performing group-specific spatial decay model for predicting water site and tap/borehole usage (Eq. 4.2). Slopes in each decay model were varied by the variable denoted in the “model” column. All models had district-specific intercepts. ELPD difference denotes the difference to the best-performing model. The *p*-value is based on a two-sided comparison (*t*-test) to the best-performing model.

4.3.9 Spatial Decay Models for Predicting Water Site and Tap/Borehole Usage

We assessed whether spatial decay models outperformed the logistic regression models in predicting the specific water sites and taps/boreholes individuals used. Using Bayesian non-linear regression models, two decay functions commonly used in gravity models—power-law decay and exponential decay [41]—as well as a Hill function were implemented. A model with exponential decay performed significantly better than one with power-law decay (difference in ELPD = -159.3, standard error (SE) = 21, $p < 0.01$). The exponential decay model also outperformed a decay model with a Hill function (difference in ELPD = -23.9, standard error (SE) = 13.8, $p = 0.08$). A visual comparison of the decay functions is provided in Fig. S4.3. The exponential decay closely matches the data at distances < 1000m, and while the Hill function provides a better fit for site usage at distances between 1000–2000m, its fatter tail leads it to

overestimate site usage at distances $> 2000\text{m}$ (Fig. S4.3). The exponential decay therefore appears to provide the best overall fit (Fig. S4.3).

The global spatial decay model (Eq. 4.1) with an exponential decay significantly outperformed the best-performing baseline model (top row of Table 4.3) based on ELPD (difference in $\text{ELPD} = -23$, $\text{SE} = 2.7$, $p < 0.01$). In particular, the spatial decay model performed better than the top baseline model in predicting water contact at water sites close to households (Fig. S4.7). The spatial decay model had an auROC of 0.882. Results suggested that for people living adjacent to water sites (20m), the probability of use was 69% (95% credible interval [CrI]: 60–78%). At 100m distance, it was estimated that 51% (95% CrI: 45–57%) used these sites. This probability was 11% (95% CrI: 10–13%) at 500m and reached 1% (95% CrI: 0.7–1.4%) at 1153m (Fig. 4.4a).

The same modelling approach was applied to tap/borehole usage, revealing a more pronounced decay than for water sites. The spatial decay model had an auROC of 0.867. For individuals adjacent to taps/boreholes, the estimated probability of any usage was 61% (95% CrI: 49–74%), 8%-points lower than for water sites. The estimated probability of use was

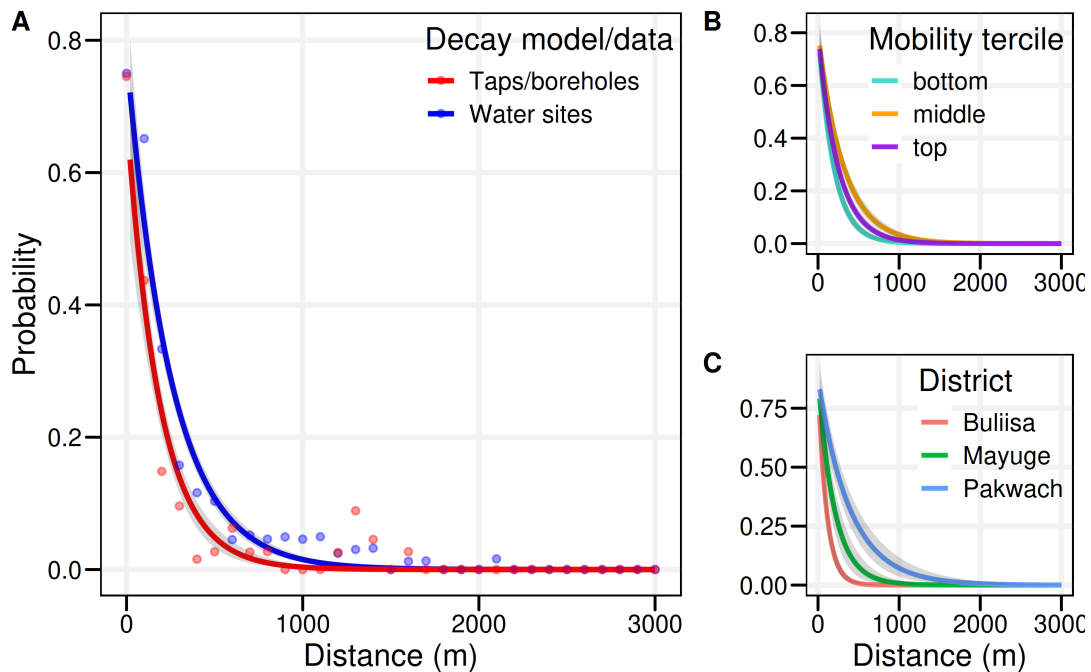


Figure 4.4. Spatial decay models for water site and tap/borehole usage over household distance. (A) Global decay models. (B) Best-performing decay model for water site usage (from Table 4.4) with separate decay functions for each mobility tercile. (C) Best-performing decay model for tap/borehole usage (from Table 4.4) with separate decay functions for each district.

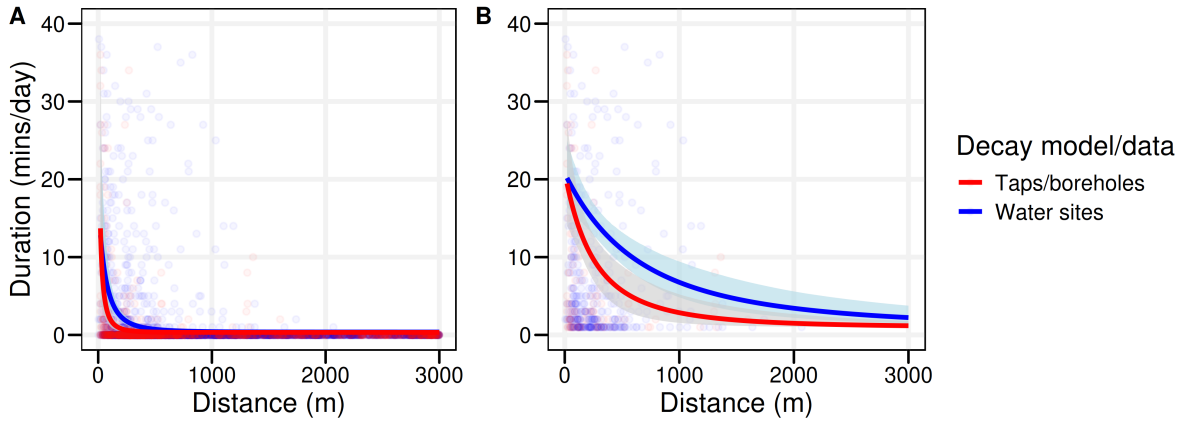


Figure 4.5. Spatial decay models for water site and tap/borehole usage duration. (A) Duration among all individuals. (B) Duration among individuals using water sites and taps/boreholes. Duration was measured as minutes per day.

40% (95% CrI 34–46%) at 100m, 5% (95% CrI: 3–6%) at 500m, and reached 1% (95% CrI: 0.5–1.7%) at 792m, 361m less than for water sites (Fig. 4.4a).

For both water site and tap/borehole models, it was investigated whether incorporating specific distance decays by gender, age, occupation, mobility, activity match, and district improved model performance (Eq. 4.2). For water sites, a model with district-specific intercepts and separate decay functions by mobility tercile performed best (Table 4.4, Fig. 4.4b). The performance of the second-best model, which included district-specific intercepts and slopes, was not significantly different based on ELPD ($p = 0.52$). For taps/boreholes, a model with district-specific intercepts and slopes significantly outperformed all other models (Table 4.4, Fig. 4.4c). Here, the model with mobility-tercile-specific decays performed third-best. A model where decays were estimated separately for taps and for boreholes performed worst. Estimated distance decays from the best-performing group-specific water site and tap/borehole usage models with a global intercept b_0 are shown in Fig. 4.4b–c. For models with district-specific intercepts, see Figs. S4.8 and S4.9.

4.3.10 Spatial Decay Models for Predicting Water Site and Tap/Borehole Usage Duration

The same general modelling framework as for any usage was applied to predict the duration of water site and tap/borehole usage. For this task, zero-inflated negative binomial models were used. The spatial decay model predicted that for individuals living directly next to water sites,

usage duration was 13 min per day (95% CrI: 6–27 min). At 100m, the estimated duration was 5 min (95% CrI: 4–8 min) and at 500m, it was 0.7 min (95% CrI: 0.6–0.9 min; Fig. 4.5a). The estimated usage duration of taps/boreholes for individuals in their direct vicinity was 14 min per day (95% CrI: 4–40 min) which was similar to water sites, but duration decayed comparatively more quickly. At 100m, the predicted tap/borehole usage duration was 2 min (95% CrI: 1–4 min), and at 500m, it was 0.4 min (95% CrI: 0.3–0.5 min, Fig. 4.5a).

Using negative binomial regression, usage duration was modelled by retaining only water sites and taps/boreholes with non-zero duration for each individual. For water sites, the average duration among site users was 20 min per day (95% CrI: 16–25 min) for those living in the direct vicinity (Fig. 4.5b). This duration reduced with distance but remained positive, with an estimated duration of 3.4 min (95% CrI: 2.2–5.6 min) at 2000m. Among users of taps/boreholes, duration decayed faster compared to users of water sites (Fig. 4.5b). Predicted taps/borehole duration was 20 min (95% CrI: 14–28 min) for individuals living next to such facilities, and 1.5 min (95% CrI: 1.2–2.7 min) at 2000m. This analysis was limited by the small number of usage events at greater distances, as only 18 and seven individuals were observed using water sites and taps/boreholes, respectively, at > 1000m from their households.

For all zero-inflated negative binomial models, it was assessed whether zero inflation was adequately modelled using the test data. For water sites, the ratio of predicted versus true durations ≤ 1 minute was 0.85. For taps/boreholes, this ratio was 0.93. Model performance was also assessed via the correlation between predicted versus actual durations (Pearson correlation coefficient [ρ] for water sites = 0.56; ρ for taps/boreholes = 0.29).

4.3.11 Spatial Decay Models Accounting for the Number of Available Water Sites

So far, spatial decay models have estimated probabilities for site usage independently of the number of sites available to an individual. However, this may not be realistic, as the number of sites in proximity to an individual is likely to influence the usage of a given site, even when holding site distance constant. For instance, if an individual lives 500m from a water site and this is their closest site, the usage probability for this individual is likely to be higher than for someone for whom the same site at 500m distance is the fifth closest site. This was tested

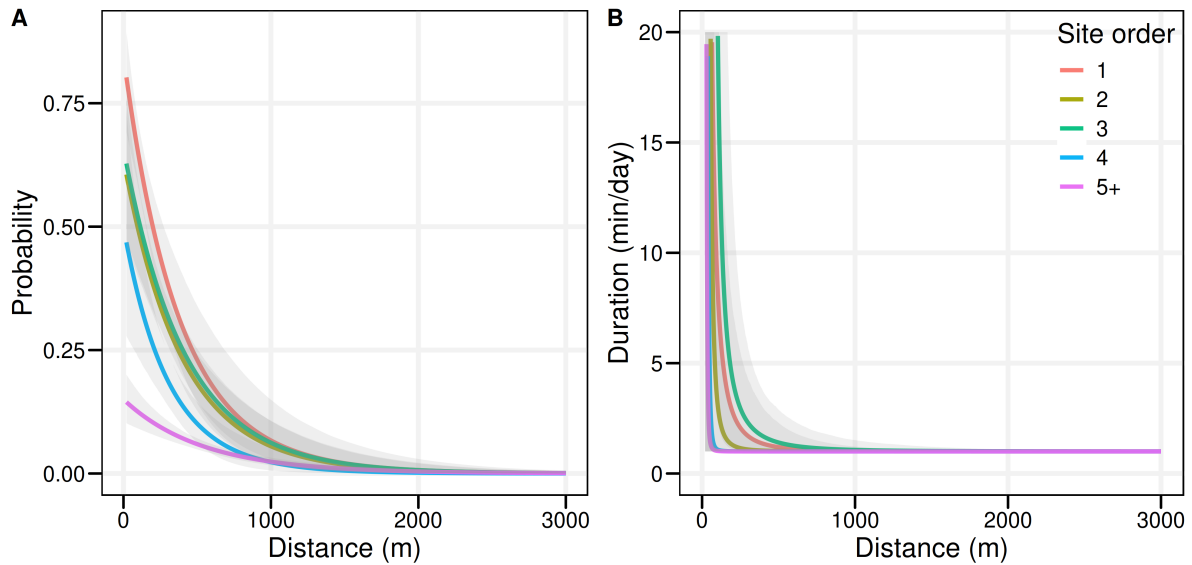


Figure 4.6. Spatial decay models accounting for the number of available water sites. (A) Spatial decay model predicting site usage probability based on whether a water site was the closest, second closest, third closest, fourth closest, or fifth closest site (or a site further away). (B) Spatial decay model predicting site usage duration based on whether a water site was the closest, second closest, third closest, fourth closest, or fifth closest site (or a site further away). The model in (B) was estimated as a zero-inflated negative binomial model.

by generating a site ordering variable and estimating separate decay functions for the closest, second closest, third closest, fourth closest, and fifth closest, or sites further away site. Fig. 4.6a shows an almost monotonic decline in site usage probability over site order. For a site at 100m household distance, the probability of usage was 65.3% (95% CI 58.6–71.6) if it was the closest site, 57.2% (95% CI 47.4–66.3) if it was the second closest site, 49.6% (95% CI 38.8–61.7) if it was the third closest site, 31.5% (95% CI 19.6–46.7) if it was the fourth closest site, and 11.8% (95% CI 8.7–15.5) if it was the fifth closest site or below (Fig. 4.6a). A similar monotonic decay, albeit somewhat steeper, was observed for site usage duration, where average duration declined as site order increased (Fig. 4.6b). For both site usage probability and duration, it appeared that usage levels were relatively similar between the first to third closest sites and substantially dropped thereafter.

4.3.12 Crowding Out of Water Site Usage by Tap/Borehole Usage

The potential crowding out effect of taps/boreholes on water site usage was investigated by comparing the performance of two modelling approaches in predicting water site usage: the

best-performing spatial decay model without crowding out (Eq. 4.2, Fig. S4.8) versus a spatial decay model with crowding out (Eq. 4.3). Using leave-one-out cross-validation, the model without crowding out performed slightly, but not significantly, better than the model with crowding out (EPLD difference = -0.9 , SE = 0.6 , $p = 0.12$). For individual-level predictions,

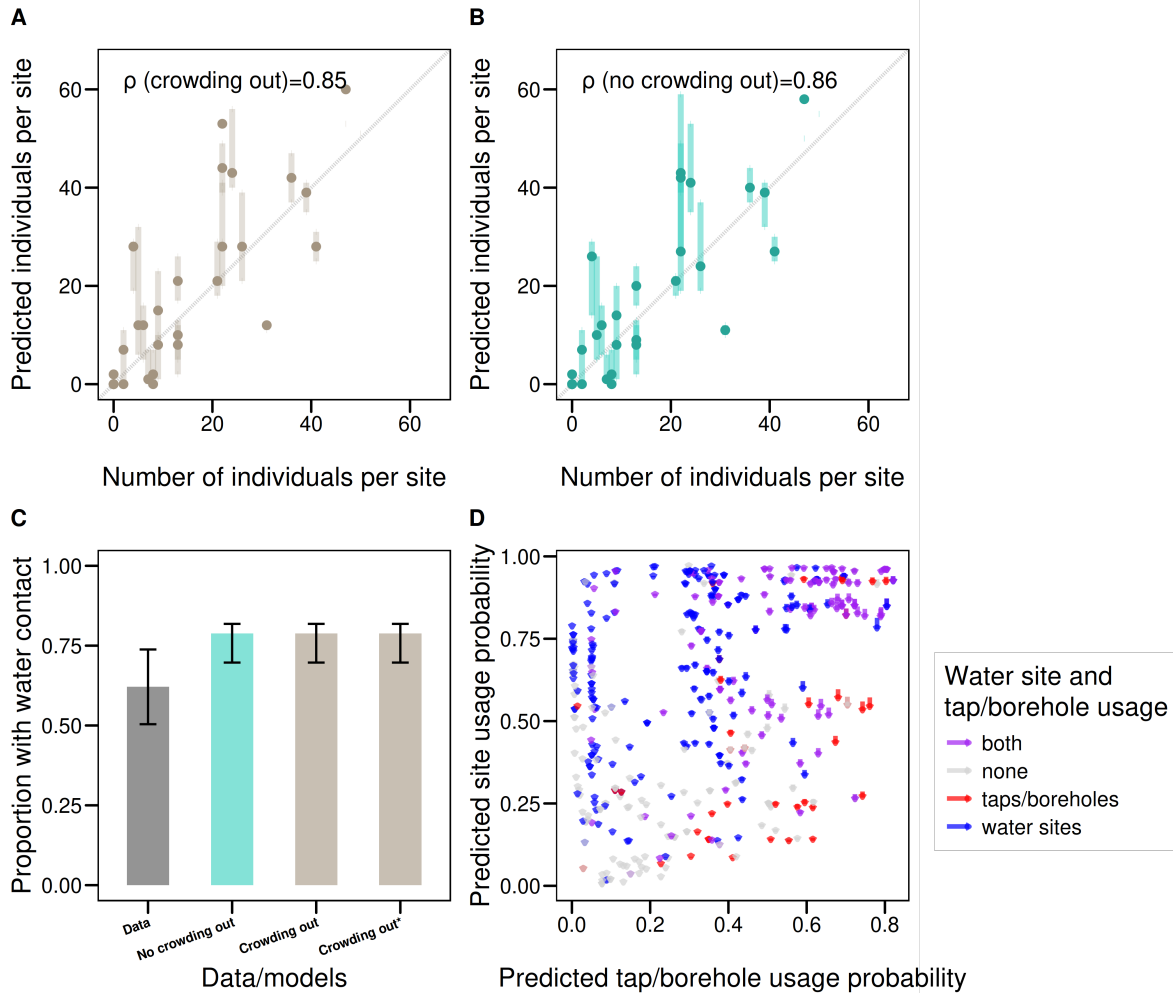


Figure 4.7. Comparison of spatial decay models with and without crowding for predicting water site usage. (A–B) Predicted number of individuals using each water site versus the true number. (C) True proportion of individuals with water contact versus the proportion predicted in spatial decay models with and without crowding out. The crowding out* model represents a model in which the proportion of individuals was estimated under a counterfactual scenario where no taps/boreholes were present. Thus, the difference between the crowding out versus crowding out* models represents the estimated impact of taps/boreholes on water site usage. (D) Scatter plot showing the predicted probability of tap/borehole usage versus the predicted probability of water site usage for each individual. The length of the arrows indicates the reductions associated with the presence of taps/boreholes for each individual. Again, this was estimated by comparing the observed values for tap/borehole coverage to a counterfactual scenario where no taps/boreholes were present. The colours in (D) represent the true usage patterns for each individual.

the performance difference was very small (model without crowding out $\text{auROC} = 0.874$ versus model with crowding out $\text{auROC} = 0.875$). At the water site level, the correlation between the predicted number of individuals using a water site versus the true number was almost identical (ρ decay model without crowding out = 0.86 and ρ decay model with crowding out = 0.85, Fig. 4.7a–b). Both the spatial decay models with and without crowding out predicted higher proportions (0.79 [95% CI: 0.70–0.82] and 0.79 [95% CI: 0.70–0.82], respectively) compared to the true proportion of individuals using water sites (0.62 [95% CI: 0.50–0.74], Fig. 4.7c).

Results from the crowding out model were used to estimate the influence of taps/boreholes on water site usage. The crowding out parameter α had a value of 0.10 (95% CrI: 0.01–0.28), suggesting that taps/boreholes could reduce the probability of water site usage by 10%. However, this magnitude would only be achieved for individuals living directly next to taps/boreholes. The overall influence of taps/boreholes was estimated by comparing the crowding out models with taps/boreholes to a counterfactual situation in which the probability of tap/borehole visits for all individuals was set to zero. Based on this comparison, the overall effect of taps/boreholes on water contact was small; only 3/452 individuals (< 1%) were predicted to have no water contact due to the presence of taps (95% CI: –31–32 individuals). On average, water contact probability was reduced by 1% due to the presence of boreholes (Fig. 4.7c).

As shown in Fig. 4.7d, these reductions were unevenly distributed across the population. Individuals with the highest predicted water site usage probability also tended to have the highest tap/borehole usage probabilities and thus experienced the largest benefits. The spatial decay models with and without crowding out produced similar positive correlations between water site usage and tap/borehole usage probabilities ($\rho = 0.21$, 95% CI: 0.12–0.29 and $\rho = 0.22$, 95% CI: 0.13–0.31, respectively) whose 95% CI contained the correlation observed in the data ($\rho = 0.15$, 95% CI: 0.06–0.24).

4.3.13 Crowding Out of Water Site Usage Duration by Tap/Borehole Usage

The crowding out analysis for water site usage was repeated for water contact duration. Here, the crowding out model was a decay model in which water contact duration was reduced by the probability of using taps/boreholes. Based on ELPD, it was found that the joint model performed

slightly, but not significantly, better than the no crowding out model (EPLD difference = -0.6 , SE = 1.5 , $p = 0.68$). The correlations between the predicted durations with the true durations

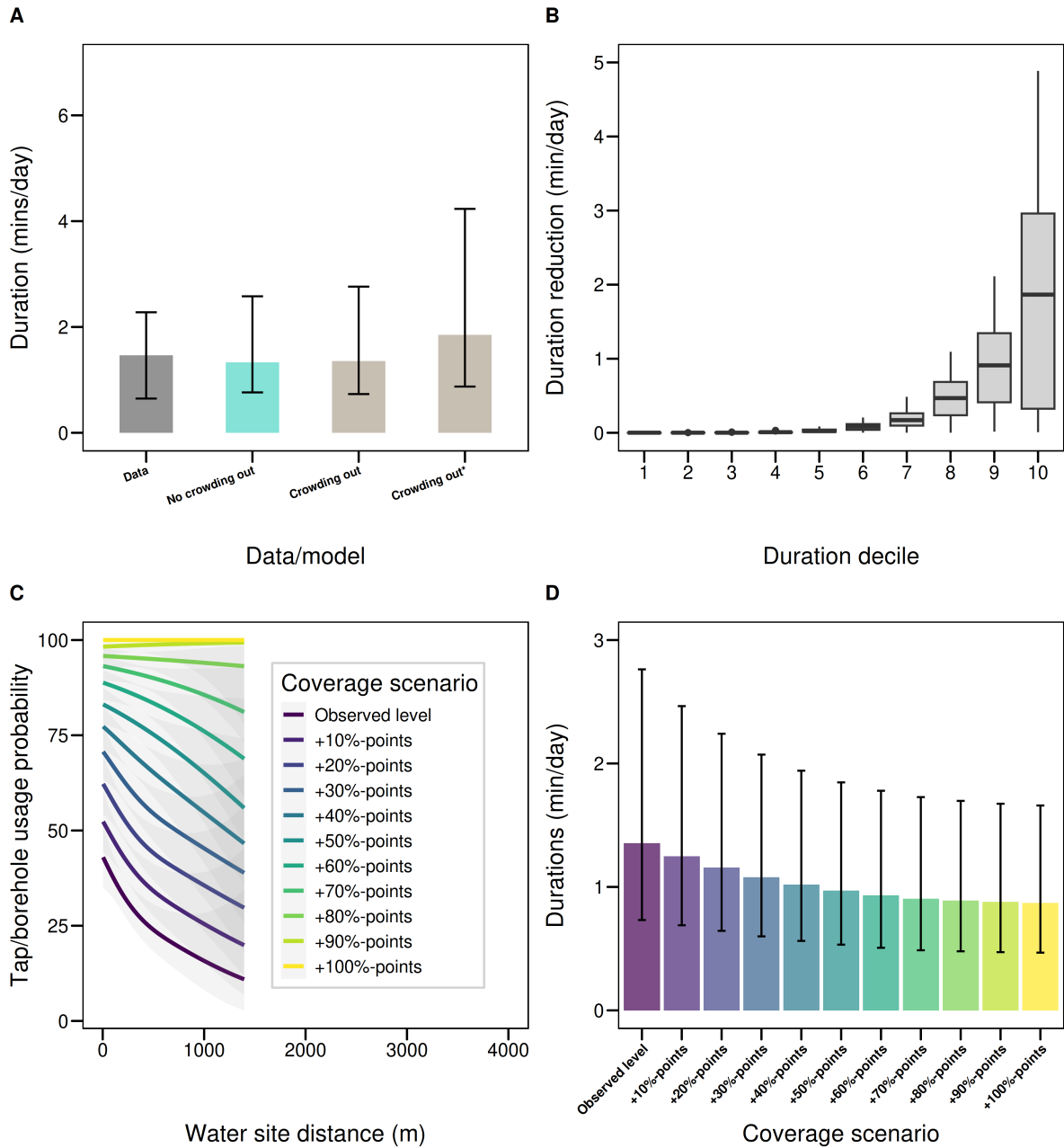


Figure 4.8. Crowding out effects of taps/boreholes on water site usage duration. (A) True duration of water contact in min per day versus estimated durations predicted using the crowding out and no crowding out models. The crowding out* model represents a model in which the water site usage duration of individuals was estimated under a counterfactual scenario where no taps/boreholes were present. (B) Reductions in water site usage duration by duration decile. (C) Tap/borehole usage probability estimated from spatial decay model, (black line) and ten different hypothetical scenarios/interventions where the estimated probability tap/borehole usage was increased by 10–100%-points for all individuals from the observed level. (D) Estimated reductions in water site usage duration under the scenarios shown in (C).

were almost identical in the models with and without crowding out ($\rho_s = 0.39$ in both cases).

A comparison of α between the crowding out models for any water site usage versus site usage duration suggested that the crowding out effect was stronger for duration than for any usage. We observed a 31% reduction in water site duration (95% CrI: 0.02–0.63) versus a 10% reduction (95% CrI: 1%–28%) in the probability of any site usage. The estimated overall effect of taps/boreholes on water site usage was moderate; without the presence of taps, water site duration was 1.8 min per person per day (95% CrI: 0.9–4.2), while with taps/boreholes, the respective duration was 1.4 min (92% CrI: 0.7–2.8). This represented a reduction of 0.4 min or 22% (Fig. 4.8b). Almost all reductions ($> 95\%$) were concentrated among the individuals in the top three declines of water site usage duration (Fig. 4.8b). We evaluated the impact of hypothetical WASH interventions in which the probability of all individuals to use taps/boreholes was increased by 10–100%-points (Fig. 4.8c). Using the spatial decay model with crowding out (Eq. 4.5), we show that reductions in water site usage from higher tap/borehole access decrease exponentially, with barely any further reductions beyond a 50%-point tap/borehole usage probability increase (Fig. 4.8d). The reason for these results is that people with the highest durations of water contact already had the highest levels of tap/borehole usage (Fig. 4.8c). Therefore, the potential to further crowd out their open water contact with safe water supplies was limited.

4.3.14 Population-Wide Predictions of Water Site Usage

Building footprints of 5177 structures in Pakwach were extracted from Google–Microsoft Open Buildings Dataset (see Fig. 4.9a for an example). Assuming that each structure represented a household, we predicted the probability of water site usage for all households (Fig. 4.9a). The model we used for this purpose has been shown to predict the true number of individuals using water sites in the test dataset with high accuracy ($\rho = 0.86$; Fig. 4.7b). Applying this model to all households in Pakwach, we estimated population-level site usage. The results suggest that there were substantial differences in site usage, with one site predicted to be used by 61 households and another by 740 households (Fig. 4.9b).

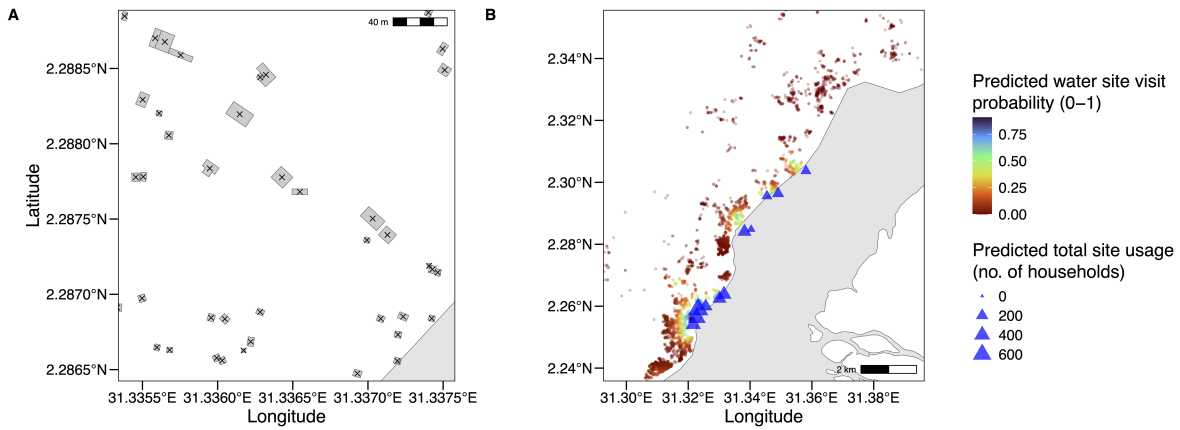


Figure 4.9. Population-wide predictions of water site usage. (A) Detail of Google–Microsoft Open Building dataset in Pakwach. Black crosses indicate the centroids of each building used to calculate the distance between every building and all local water sites. (B) Population-wide predictions of the water site usage probability for each household based on 5177 buildings in Pakwach. Blue triangles show all local water sites mapped by the SchistoTrack study, and their respective sizes indicate the number of households predicted to use each site (multiple assignments possible).

4.3.15 Minimum Data Needs

The minimal number of individuals and observation days needed for the spatial decay models was established by holding out individuals, entire GPS logger days, and districts. When using the entire dataset of sampled individuals (mean number of GPS observation days = 11), the performance of the model plateaued after just 15 individuals were used to model the decay function (Fig. 4.10a). Using a random subset of all days (1–6 days) revealed that even with only one day of data, model performance for predicting water site usage was similar to when all six days were used (Fig. 4.10b). Holding out entire districts revealed that the data from Buliisa was best able to predict the decay function for the two held-out districts (auROC = 0.910, 95% CI: 0.905–0.916), Mayuge and Buliisa had similar performance for predicting the held-out districts (auROC Mayuge = 0.859, 95% CI: 0.853–0.866 and auROC Pakwach = 0.859, 95% CI: 0.851–0.866). As the decay functions of Mayuge and Pakwach were more similar (Fig. S4.8), training the model on Mayuge performed best at predicting Pakwach, and vice versa (Fig. 4.10c).

4.3.16 Robustness to Sampling Frequency

We also assessed the impact of GPS logger sampling frequency (2 min) on the parameter estimates in the global spatial decay model predicting water site usage (Fig. S4.10). The

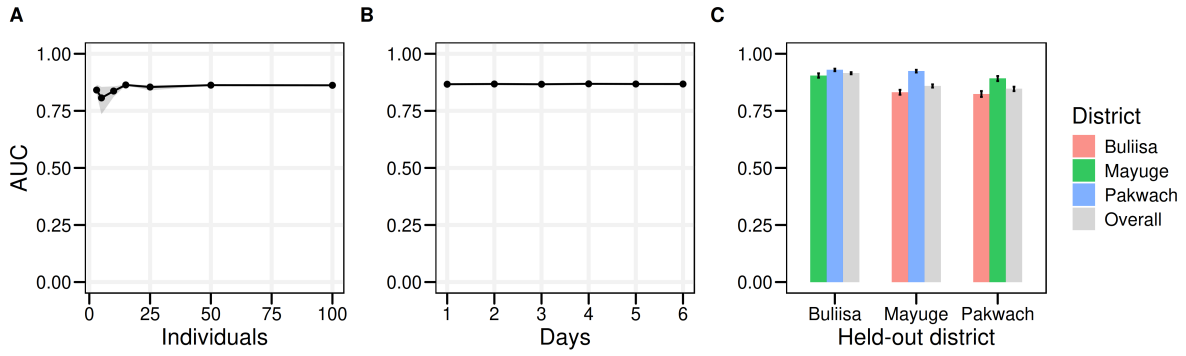


Figure 4.10. Minimum data needs for estimating the spatial decay of water site usage. (A) auROC by number of individuals used to estimate the spatial decay. (B) auROC by number of GPS logger days used to estimate the spatial decay. (C) Evaluating performance on held-out districts. For (A–B) estimation was performed by randomly sampling from the training data three times and evaluating the auROC on held-out test data. For (C), models were estimated on the full data from one district and performance was evaluated separately for each of the held-out districts and for both held-out districts jointly (overall columns). The x -axis label in (C) indicates the held-out district.

results are largely robust to removing low-duration events. For the intercept b_0 , we found that the estimated credible interval contains the ground truth value until all events < 5 min were removed (Fig. S4.10a). For the exponential decay parameter b_1 , we found that even when removing all events < 10 min, the estimated credible interval still contains the ground truth value of b_1 (Fig. S4.10b).

4.3.17 Representativeness

Sufficient data from GPS loggers was retrieved for 452/584 participants. To investigate differences between individuals with and without missing data, their demographic characteristics were compared in Table 4.5. Additionally, comparisons were made between individuals selected for the GPS logger study ($n = 584$) and SchistoTrack participants from the same villages who were not selected for the GPS logger study ($n = 379$). Participants with sufficient GPS data were significantly more likely to be male (50.0%) compared to those with missing data (40.9%, $p = 0.01$). There was a relationship between gender and missingness because both gender distribution and missingness varied across districts. In Mayuge where 65.2% (86/132) of all missing participants were located, 39.4% (126/320) of all SchistoTrack participants in the GPS logger study villages were male, whereas in Pakwach, the district with the lowest missingness, 53.4% were male (149/320). Participants with GPS data did not significantly differ from those with

Table 4.5. Comparison of participants with sufficient GPS data versus those with missing data and the wider SchistoTrack study population.

Variable	GPS complete (n= 452)	GPS missing (n= 132)	Not selected for GPS (n= 379)	P-value
Gender = male (%)	226 (50.0)	54 (40.9)	149 (39.3)	0.01
Age $\geq$ 15 years (%)	254 (56.2)	75 (56.8)	214 (56.5)	0.99
Water contact = yes (%)	231 (51.1)	58 (43.9)	169 (44.6)	0.12
Occupation (%)				0.19
none	259 (57.3)	84 (63.6)	225 (59.4)	
fishing	41 (9.1)	10 (7.6)	24 (6.3)	
fishmongering	22 (4.9)	2 (1.5)	19 (5.0)	
farming	93 (20.6)	31 (23.5)	73 (19.3)	
other	37 (8.2)	5 (3.8)	38 (10.0)	
Distance to the closest water site (m) (Mean $\pm$ SD)	278.6 $\pm$ 270.0	298.0 $\pm$ 198.4	289.1 $\pm$ 256.3	0.65
Household drinking water source = improved (%)	226 (50.0)	33 (25.0)	160 (42.2)	<0.01
District (%)				<0.01
Mayuge	105 (23.2)	86 (65.2)	129 (34.0)	
Buliisa	164 (36.3)	27 (20.5)	132 (34.8)	
Pakwach	183 (40.5)	19 (14.4)	118 (31.1)	

Notes: For binary variables, number and proportions [N (%)] are shown and Chi-squared tests were used to test for group differences. For all normally distributed continuous variables, mean and standard deviation (mean $\pm$ sd) are shown and two sample *t*-tests were performed to test for group differences. For all non-normally distributed variables, median and interquartile range (median [IQR]) are shown and Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

missing data or the broader SchistoTrack study population in terms of age, self-reported water contact, occupation, and distance to the closest water site (Table 4.5). However, participants with GPS data reported higher use of an improved household drinking water source compared to those with missing data (50.0% versus 25.0%, $p < 0.01$).

4.3.18 Wear Non-Compliance

As the mean number of GPS observation days varied by district (Pakwach = 13, Buliisa = 12, Mayuge = 11), wear compliance was investigated using district-specific terciles of the number of GPS points per person per day (Table S4.2). This allowed us to understand whether relative levels of compliance varied based on socio-demographic characteristics. We found no significant differences across any of the included variables, except for age. The proportion of those aged

≥ 15 years was higher (64.7%) in the tercile with the highest number of points compared to the bottom tercile (51.0%, p -value = 0.04), indicating that adults may have been more compliant. There was no evidence that device failure due to water contact introduced information bias as the number of observation days was similar between participants with and without GPS-derived water contact (11.2 versus 11.0 days, $p = 0.72$).

4.4 Discussion

Currently, strategies are lacking for implementing the 2022 WHO schistosomiasis control guideline which recommends improved safe water supplies and the reduction of contact with infected water sources for long-term schistosomiasis control and prevention [6, 52]. This chapter presented evidence regarding the extent to which the presence of public safe water infrastructure, specifically taps/boreholes, could crowd out the use of infested freshwater sources. Using data from wearable GPS loggers on 452 participants across 12 rural villages in Uganda, we derived measures of contact with open water sites and taps/boreholes. We built statistical models to predict which sites and taps/boreholes were likely to be used. The results suggest that the presence of taps/boreholes can potentially reduce the probability of individuals to contact water sites by 10% and the duration of water site usage by 31%.

We demonstrate that spatial decay models provide a useful extension of gravity models for modelling individual-level behaviours. Gravity models incorporating spatial decay functions have primarily been applied with aggregate human mobility data to predict population-level movements or disease transmission [26, 29, 41]. Here, simple spatial decay models were demonstrated to accurately predict individual-level water contact on small spatial scales, despite the considerable complexity of human behaviour. A global decay model with household distance to water sites as the sole covariate achieved an auROC of 0.882. When using spatial decay models trained on one district at a time to predict water site usage in held-out districts across Uganda, we found that model auROCs were > 0.8 across held-out districts. More complex spatial decay models with varying decays by district or mobility were not able to exceed the model performance of the global model on unseen test data. Importantly, we demonstrate that a global decay model requires very limited data inputs. Sensitivity analyses revealed that GPS

data from 15 individuals over ten observation days would be sufficient to estimate the spatial decay function of water site usage. Besides the GPS data, locations of all local water sites and participant households would need to be mapped using handheld GPS devices or tablets. Publicly available building footprint data or population data from WorldPop [51, 53] could be utilised with our models to obtain population-wide predictions of water site usage. As spatial decay models require limited field data collection and were shown to be generalisable across different settings in Uganda, they may represent a versatile tool for obtaining large-scale water contact information.

This chapter focused on predicting which specific water sites and taps/boreholes people used and found that household distance and mobility, not demographics, were the major influencing factors. This highlights the importance of spatial proximity and movement patterns in guiding the choice of water sites and taps/boreholes visited. The relationships between both distance and mobility with water site usage were non-linear. Water site and tap/borehole usage decayed exponentially with very low usage levels observed beyond 1-kilometre distance from the household. Usage of taps/boreholes decayed faster with distance than usage of water sites. For human mobility, we found that it was non-monotonically related to water site and tap/borehole usage patterns; individuals in the lowest mobility tercile used water sites and taps/boreholes closer to their household locations compared to those in the middle tercile. However, this relationship did not hold for the highest mobility tercile where greater distances were only observed for ages 20–40, coinciding with the years of highest water contact (see Chapter 3).

The ability of spatial decay models in accurately predicting tap/borehole usage (auROC of 0.867) enabled us to devise a decay model with crowding out to estimate reductions in water site usage due to tap/borehole presence. Our results suggest that taps/boreholes reduce open water contact by 10%. However, this magnitude is mostly hypothetical as it would materialise only if taps/boreholes were built next to the household. The estimated overall effect of taps/boreholes on water site usage was small ($< 1\%$ reduction) as individuals lived a median of 251m from the closest tap/borehole. Our findings suggest that taps/boreholes primarily reduce the duration of open water contact rather than eliminating water site usage completely. The spatial decay model with crowding out estimated that the presence of taps/boreholes may reduce usage duration of

water sites by up to 31%. At the observed coverage levels, it was estimated that taps/boreholes decreased the average open water contact usage duration of individuals from 1.8 min per day to 1.4 min, a 22% reduction. The effect sizes observed here may be taken to represent an upper limit of crowding out effects. Participants in the GPS sample lived close to water sites and had high levels of tap/borehole usage. This should have enabled us to observe large crowding out effects if they existed, but the effects we found were moderate. In the Bayesian model selection, spatial decay models with crowding out did not significantly outperform models without crowding out, indicating that observed water contact patterns could be successfully predicted without incorporating information on safe water supplies. Despite the moderate crowding effects, the results may still be encouraging, particularly considering that a series of randomised controlled trials—the WASH benefits and SHINE trials which were conducted across Bangladesh, Kenya, Mali, and Zambia—found no positive effects of WASH interventions on clinical outcomes [54–57]. In future trials, GPS data collection could be used to obtain information on water contact, latrine, and tap/borehole use. By overcoming reporting and social desirability biases in self-reported data, GPS data could help to understand whether WASH interventions achieve the desired behaviour change.

By identifying at-risk populations for open water contact at fine spatial scales, decay models could be deployed as support tools in national control programmes. There has been an increasing emphasis on precision mapping, i.e. mapping infection prevalence at low administrative levels, such as the sub-district level, to support more appropriate MDA treatment strategies that minimise over- and under-treatment [58–60]. However, interim results from the Schistosomiasis Oversampling Study by the Coalition for Operational Research on Neglected Tropical Diseases highlighted that heterogeneity can extend beyond the sub-district level [61]. The working group found that satellite and environmental data were not necessarily effective for finding at-risk areas and suggested collecting data on local fishing activities and recreational water contact of children as proxies for local risk [61]. To obtain water contact information, especially for large-scale programmes, the cost and data collection efforts required for using decay models with GPS loggers could possibly be lower than conducting water contact surveys. Our results suggest that GPS logger studies would need a minimum of only 15 people to predict water con-

tact patterns. Beyond MDA, spatial decay models could guide implementation of multifaceted control interventions. Recommendation five of the 2022 WHO schistosomiasis control guideline recommends the deployment of “WASH interventions, environmental interventions (water engineering and focal snail control with molluscicides), and behavioural change interventions as essential measures” alongside MDA to reduce transmission [6]. For mollusciciding, the WHO previously outlined three strategies: area control, where all local waterbodies are treated; focal control, where treatment is limited to waterbodies commonly used for human water contact activities; and radius control, where all waterbodies within 500m of human settlements are treated [62]. Population-wide spatial decay models can support effective deployment of focal and radius control by identifying the most widely used water sites. Although the WHO guideline suggests 500m as the threshold for radius control, the optimal distance threshold may depend on the setting and resources available. We estimated that in SchistoTrack, the probability of individuals using water sites at 500m household distance was still 11%, suggesting that 500m may be too low of a threshold to capture all water contacts. Spatial decay models can be used as decision support tools to guide locally appropriate treatment radii. Similarly, decay models could identify population subsets with the highest water contact for targeting with behaviour change strategies.

Most prior studies on mobility within the context of schistosomiasis have leveraged mathematical models to demonstrate the role of human movement in long-range transmission [23, 24]. Proxies such as building footprints or call detail records data were previously used to infer mobility patterns [63, 64]. However, in rural African settings, limited cell tower infrastructure constrains spatial resolution to kilometres. This study demonstrates the value of GPS data for understanding mobility and small-scale exposure behaviour for environmentally transmitted diseases.

4.4.1 Limitations

There are several limitations of this chapter. We focused on water contact at sites mapped as part of the SchistoTrack study. Despite efforts to comprehensively map all water sites used by individuals of the study villages, some may have not been captured. Additionally, deriving

reliable water contact frequency measures from GPS data remains challenging. We used a relatively small buffer of 30m to avoid overestimation of water contact duration, but with such a site buffer, during an individual's water contact, some GPS points could fall outside the buffer. This makes it challenging to segment data into distinct events for counting water contact frequency. As the GPS loggers were motion activated to save battery, we were unable to distinguish between periods of immobility versus wear non-compliance of participants. This study prioritised water contact duration over frequency, as duration better reflects the intensity of water contact. More sophisticated trajectory analysis could try to derive water contact frequency measures from GPS data in the future. Another limitation is that we focused exclusively on water contacts at water sites. Consequently, we found that individuals with fishing occupation had lower water contact durations compared to those engaging in fishmongering as fishermen spend extended periods of time away from the shoreline, which is not captured as water contact in this study. By contrast, fishmongers have high durations of water contact because they typically wash and prepare fish at water sites. The focus on water sites was motivated by their relevance for infection as the risk of cercarial exposure is highest in shallow water while contact with deeper water ($> 10\text{m}$) may represent less of a risk as snail density is low in these zones [65]. Finally, this study focused on spatial dependencies in human behaviour but did not investigate the temporal dynamics. Over the two weeks of the study, few individuals had clear regular site and tap/borehole visitation patterns which could have been amenable to spatio-temporal modelling. Due to the small-scale nature of this study, we did not include climate or hydrological variables, as there would have been insufficient variation in these patterns to make meaningful inferences about how they influence water contact patterns. Future work conducted at larger geographic scales where GPS data are collected via smartphones to enable longer tracking periods could help to understand monthly or seasonal water contact patterns and the influence of environmental variability. For our purpose of building a spatial decay model, we showed that with 452 individuals, even a single day of GPS data was sufficient to capture the spatial dynamics of water contact at a single time point.

4.4.2 Conclusion

This chapter underscores the importance of household distance and mobility for determining water site and tap/borehole usage patterns. Using GPS data from 452 participants in Uganda, spatial decay models were shown to effectively predict individual water site and tap/borehole visitation behaviour. While communities need to be involved in WASH planning to ensure uptake and usage of new infrastructure, spatial decay models offer practical applications for informing optimised WASH infrastructure placement, to reduce open water contact. The results highlight the need for multifaceted, spatially informed strategies to address schistosomiasis transmission risk.

References

1. United Nations. Transforming our World: The 2030 Agenda for Sustainable Development. New York, NY: United Nations, 2015
2. UNICEF. Drinking water, sanitation and hygiene in households by country, 2000-2022. <https://data.unicef.org/topic/water-and-sanitation/drinking-water/>. [Accessed on: 2024 Dec 31]
3. Prüss-Ustün A, Wolf J, Bartram J, Clasen T, Cumming O, Freeman MC, Gordon B, Hunter PR, Medlicott K, and Johnston R. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health* 2019 Jun; 222:765–77. DOI: 10.1016/j.ijheh.2019.05.004
4. Wolf J, Johnston RB, Ambelu A, Arnold BF, Bain R, Brauer M, Brown J, Caruso BA, Clasen T, Colford JM, Mills JE, Evans B, Freeman MC, Gordon B, Kang G, Lanata CF, Medlicott KO, Prüss-Ustün A, Troeger C, Boisson S, and Cumming O. Burden of disease attributable to unsafe drinking water, sanitation, and hygiene in domestic settings: a global analysis for selected adverse health outcomes. *The Lancet* 2023 Jun; 401:2060–71. DOI: 10.1016/s0140-6736(23)00458-0
5. Grimes JE, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasites & Vectors* 2015 Dec; 8:156. DOI: 10.1186/s13071-015-0766-9
6. World Health Organization. WHO guideline on control and elimination of human schistosomiasis. Geneva: World Health Organization, 2022
7. Prüss-Ustün A, Wolf J, Corvalán C, Bos R, and Neira M. Preventing disease through healthy environments: A global assessment of the burden of disease from environmental risks. 2016. Tech. rep. World Health Organization, 2020
8. Grimes JET, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The Relationship between Water, Sanitation and Schistosomiasis: A Systematic Review and Meta-analysis. *PLoS Neglected Tropical Diseases* 2014 Dec; 8:e3296. DOI: 10.1371/journal.pntd.0003296
9. Grimes JET, Tadesse G, Mekete K, Wuletaw Y, Gebretsadik A, French MD, Harrison WE, Drake LJ, Gardiner IA, Yard E, and Templeton MR. School Water, Sanitation, and Hygiene, Soil-Transmitted Helminths, and Schistosomes: National Mapping in Ethiopia. *PLoS Neglected Tropical Diseases* 2016 Mar; 10:e0004515. DOI: 10.1371/journal.pntd.0004515
10. Knopp S, Person B, Ame SM, Ali SM, Hattendorf J, Juma S, Muhsin J, Khamis IS, Mohammed KA, Utzinger J, Hollenberg E, Kabole F, and Rollinson D. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. *The Lancet Global Health* 2019 Aug; 7:e1118–e1129. DOI: 10.1016/s2214-109x(19)30189-5

11. Fulford A, Webster M, Ouma J, Kimani G, and Dunne D. Puberty and Age-related Changes in Susceptibility to Schistosome Infection. *Parasitology Today* 1998 Jan; 14:23–6. doi: 10.1016/s0169-4758(97)01168-x
12. Ouma J, Fulford A, Kariuki H, Kimani G, Sturrock R, Muchemi G, Butterworth A, and Dunne D. The development of schistosomiasis mansoni in an immunologically naive immigrant population in Masongaleni, Kenya. *Parasitology* 1998 Aug; 117:123–32. doi: 10.1017/s0031182098002935
13. Gryseels B. Human resistance to *Schistosoma* infections: Age or experience? *Parasitology Today* 1994 Jan; 10:380–4. doi: 10.1016/0169-4758(94)90226-7
14. Vaz Nery S, Traub RJ, McCarthy JS, Clarke NE, Amaral S, Llewellyn S, Weking E, Richardson A, Campbell SJ, Gray DJ, Vallely AJ, Williams GM, Andrews RM, and Clements ACA. WASH for WORMS: A Cluster-Randomized Controlled Trial of the Impact of a Community Integrated Water, Sanitation, and Hygiene and Deworming Intervention on Soil-Transmitted Helminth Infections. *The American Journal of Tropical Medicine and Hygiene* 2019 Mar; 100:750–61. doi: 10.4269/ajtmh.18-0705
15. Stewart CP, Arnold CD, Williams AM, Arnold BF, Pickering AJ, Dentz H, Kiprotich M, Lin A, Null C, Colford JM, and Dewey KG. Social Desirability Bias in a Randomized Controlled Trial That Included Breastfeeding Promotion in Western Kenya. *Current Developments in Nutrition* 2024 Sep :103779. doi: 10.1016/j.cdnut.2024.103779
16. Seto EYW, Sousa-Figueiredo JC, Betson M, Byalero C, Kabatereine NB, and Stothard JR. 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. *Geospatial health* 2012 Nov; 7:1. doi: 10.4081/gh.2012.99
17. Macklin G, Stanton MC, Tchuem-Tchuente LA, and Stothard JR. A pilot study using wearable global positioning system data loggers to compare water contact levels: *Schistosoma haematobium* infection in pre-school-age children (PSAC) and their mothers at Barombi Kotto, Cameroon. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 2018 Jul; 112:361–5. doi: 10.1093/trstmh/try059
18. Eyre M, Stanton M, Macklin G, Bartoníček Z, O'Halloran L, Eloundou Ombede D, Chuinteu G, Stewart M, LaCourse E, Tchuem Tchuente L, and Stothard J. 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 Tropica* 2020 Dec; 212:105646. doi: 10.1016/j.actatropica.2020.105646
19. Wood CL, Sokolow SH, Jones IJ, Chamberlin AJ, Lafferty KD, Kuris AM, Jocque M, Hopkins S, Adams G, Buck JC, Lund AJ, Garcia-Vedrenne AE, Fiorenza E, Rohr JR, Allan F, Webster B, Rabone M, Webster JP, Bandagny L, Ndione R, Senghor S, Schacht AM, Jouanard N, Riveau G, and De Leo GA. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. *Proceedings of the National Academy of Sciences* 2019 Oct; 116:23182–91. doi: 10.1073/pnas.1903698116
20. Lamberti O, Kabatereine NB, Tukahebwa EM, and Chami GF. *Schistosoma mansoni* infection risk for school-aged children clusters within households and is modified by distance to freshwater bodies. *PLoS ONE* 2021 Nov; 16:e0258915. doi: 10.1371/journal.pone.0258915
21. Lund AJ, Sokolow SH, Jones IJ, Wood CL, Ali S, Chamberlin A, Sy AB, Sam MM, Jouanard N, Schacht AM, Senghor S, Fall A, Ndione R, Riveau G, De Leo GA, and Lopez-Carr D. Exposure, hazard, and vulnerability and their contribution to *Schistosoma haematobium* re-infection in northern Senegal. *The Lancet Planetary Health* 2021 Apr; 5:S10. doi: 10.1016/S2542-5196(21)00094-2
22. Reitzug F, Kabatereine NB, Byaruhanga AM, Besigye F, Nabatte B, and Chami GF. Current water contact and *Schistosoma mansoni* infection have distinct determinants: a data-driven population-based study in rural Uganda. *Nature Communications* 2024 Nov; 15. doi: 10.1038/s41467-024-53519-4
23. Perez-Saez J, Mari L, Bertuzzo E, Casagrandi R, Sokolow SH, De Leo GA, Mande T, Ceperley N, Froehlich JM, Sou M, Karambiri H, Yacouba H, Maiga A, Gatto M, and Rinaldo A. 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 Neglected Tropical Diseases* 2015 Oct; 9. Ed. by Liang S:e0004127. doi: 10.1371/journal.pntd.0004127
24. Hardwick RJ, Vegvari C, Collyer B, Truscott JE, and Anderson RM. Spatial scales in human movement between reservoirs of infection. *Journal of Theoretical Biology* 2021 Sep; 524:110726. doi: 10.1016/j.jtbi.2021.110726

25. Collyer BS, Truscott JE, Mwandawiro CS, Njenga SM, and Anderson RM. How important is the spatial movement of people in attempts to eliminate the transmission of human helminth infections by mass drug administration? *Philosophical Transactions of the Royal Society B: Biological Sciences* 2023 Aug; 378. doi: 10.1098/rstb.2022.0273
26. Giles JR, Erbach-Schoenberg E zu, Tatem AJ, Gardner L, Bjørnstad ON, Metcalf C, and Wesolowski A. The duration of travel impacts the spatial dynamics of infectious diseases. *Proceedings of the National Academy of Sciences* 2020; 117:22572–9
27. Gibbs H, Musah A, Seidu O, Ampofo W, Asiedu-Bekoe F, Gray J, Adewole WA, Cheshire J, Marks M, and Eggo RM. Call detail record aggregation methodology impacts infectious disease models informed by human mobility. *PLOS Computational Biology* 2023 Aug; 19:e1011368. doi: 10.1371/journal.pcbi.1011368
28. Iyaniwura SA, Ringa N, Adu PA, Mak S, Janjua NZ, Irvine MA, and Otterstatter M. Understanding the impact of mobility on COVID-19 spread: A hybrid gravity-metapopulation model of COVID-19. *PLOS Computational Biology* 2023 May; 19:e1011123. doi: 10.1371/journal.pcbi.1011123
29. Kraemer MU, Golding N, Bisanzio D, Bhatt S, Pigott DM, Ray S, Brady O, Brownstein J, Faria N, Cummings D, et al. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific reports* 2019; 9:5151
30. Meredith HR, Giles JR, Perez-Saez J, Mande T, Rinaldo A, Mutembo S, Kabalo EN, Makungo K, Buckee CO, Tatem AJ, Metcalf CJE, and Wesolowski A. Characterizing human mobility patterns in rural settings of sub-Saharan Africa. *eLife* 2021 Sep; 10:e68441. doi: 10.7554/eLife.68441
31. Sedgwick P. The Hawthorne effect. *BMJ* 2011 Jan; 344:d8262–d8262. doi: 10.1136/bmj.d8262
32. Fleming CH and Calabrese JM. ctm: Continuous-Time Movement Modeling. 2023
33. Morris G and Conner LM. 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* 2017; 12:e0189020. doi: 10.1371/journal.pone.0189020
34. Berger-Tal O and Bar-David S. Recursive movement patterns: review and synthesis across species. *Ecosphere* 2015 Sep; 6:1–12. doi: 10.1890/es15-00106.1
35. Seidel DP, Dougherty E, Carlson C, and Getz WM. Ecological metrics and methods for GPS movement data. *International Journal of Geographical Information Science* 2018 Jul; 32:2272–93. doi: 10.1080/13658816.2018.1498097
36. Bracis C. recurse: Computes Revisitation Metrics for Trajectory Data. [https://CRAN.R-project.org: Comprehensive R Archive Network \(CRAN\),](https://CRAN.R-project.org: Comprehensive R Archive Network (CRAN),) 2024. doi: 10.32614/CRAN.package.recurse
37. Sow S, Vlas SJ de, Stelma F, Vereecken K, Gryseels B, and Polman K. The contribution of water contact behavior to the high *Schistosoma mansoni* Infection rates observed in the Senegal River Basin. *BMC Infectious Diseases* 2011 Dec; 11:198. doi: 10.1186/1471-2334-11-198
38. Hahsler M, Piekenbrock M, and Doran D. dbscan: Fast Density-Based Clustering with R. *Journal of Statistical Software* 2019; 91. doi: 10.18637/jss.v091.i01
39. Kwiringira J, Atekyereza P, Niwagaba C, and Günther I. Gender variations in access, choice to use and cleaning of shared latrines; experiences from Kampala Slums, Uganda. *BMC Public Health* 2014 Nov; 14. doi: 10.1186/1471-2458-14-1180
40. Tessema RA. Assessment of the implementation of community-led total sanitation, hygiene, and associated factors in Diretiyara district, Eastern Ethiopia. *PLoS ONE* 2017 Apr; 12. Ed. by Biswas A:e0175233. doi: 10.1371/journal.pone.0175233
41. Simini F, González MC, Maritan A, and Barabási AL. A universal model for mobility and migration patterns. *Nature* 2012; 484:96–100
42. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, and Tomasini M. Human mobility: Models and applications. *Physics Reports* 2018 Mar; 734:1–74. doi: 10.1016/j.physrep.2018.01.001
43. Arnold BF, Kanyi H, Njenga SM, Rawago FO, Priest JW, Secor WE, Lammie PJ, Won KY, and Odiere MR. Fine-scale heterogeneity in *Schistosoma mansoni* force of infection measured through antibody response. *Proceedings of the National Academy of Sciences* 2020 Aug; 117:23174–81. doi: 10.1073/pnas.2008951117

44. Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 2017; 80. doi: 10.18637/jss.v080.i01
45. Hoffman MD, Gelman A, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* 2014; 15:1593–623
46. Vehtari A, Gelman A, and Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 2016 Aug; 27:1413–32. doi: 10.1007/s11222-016-9696-4
47. Thiele C and Hirschfeld G. cutpointnr: Improved Estimation and Validation of Optimal Cutpoints in R. *Journal of Statistical Software* 2021; 98:1–27. doi: 10.18637/jss.v098.i11
48. Yang Y, Herrera C, Eagle N, and González MC. Limits of Predictability in Commuting Flows in the Absence of Data for Calibration. *Scientific Reports* 2014 Jul; 4. doi: 10.1038/srep05662
49. Masucci AP, Serras J, Johansson A, and Batty M. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E* 2013 Aug; 88. doi: 10.1103/physreve.88.022812
50. Li W, Wang Q, Liu Y, Small ML, and Gao J. A spatiotemporal decay model of human mobility when facing large-scale crises. *Proceedings of the National Academy of Sciences* 2022 Aug; 119. doi: 10.1073/pnas.2203042119
51. Google and Microsoft. Google-Microsoft Open Buildings - combined by VIDA. <https://beta.source-coop/repositories/vida/google-microsoft-open-buildings>. [Accessed on: 2024 Dec 30]
52. Lo NC, Bezerra FSM, Colley DG, Fleming FM, Homeida M, Kabatereine N, Kabole FM, King CH, Mafe MA, Midzi N, Mutapi F, Mwanga JR, Ramzy RMR, Satrija F, Stothard JR, Traoré MS, Webster JP, Utzinger J, Zhou XN, Danso-Appiah A, Eusebi P, Loker ES, Obonyo CO, Quansah R, Liang S, Vaillant M, Murad MH, Hagan P, and Garba A. Review of 2022 WHO guidelines on the control and elimination of schistosomiasis. *The Lancet Infectious Diseases* 2022 Nov; 22:e327–e335. doi: 10.1016/s1473-3099(22)00221-3
53. Tatem AJ. WorldPop, open data for spatial demography. *Scientific Data* 2017 Jan; 4. doi: 10.1038/sdata.2017.4
54. Pickering AJ, Djebbari H, Lopez C, Coulibaly M, and Alzua ML. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *The Lancet Global Health* 2015 Nov; 3:e701–e711. doi: 10.1016/s2214-109x(15)00144-8
55. Luby SP, Rahman M, Arnold BF, Unicomb L, Ashraf S, Winch PJ, Stewart CP, Begum F, Hussain F, Benjamin-Chung J, Leontsini E, Naser AM, Parvez SM, Hubbard AE, Lin A, Nizame FA, Jannat K, Ercumen A, Ram PK, Das KK, Abedin J, Clasen TF, Dewey KG, Fernald LC, Null C, Ahmed T, and Colford JM. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. *The Lancet Global Health* 2018 Mar; 6:e302–e315. doi: 10.1016/s2214-109x(17)30490-4
56. Null C, Stewart CP, Pickering AJ, Dentz HN, Arnold BF, Arnold CD, Benjamin-Chung J, Clasen T, Dewey KG, Fernald LCH, Hubbard AE, Kariger P, Lin A, Luby SP, Mertens A, Njenga SM, Nyambane G, Ram PK, and Colford JM. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. *The Lancet Global Health* 2018 Mar; 6:e316–e329. doi: 10.1016/s2214-109x(18)30005-6
57. Humphrey JH et al. Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *The Lancet Global Health* 2019 Jan; 7:e132–e147. doi: 10.1016/s2214-109x(18)30374-7
58. Tchuem Tchuente LA, Stothard JR, Rollinson D, and Reinhard-Rupp J. Precision mapping: An innovative tool and way forward to shrink the map, better target interventions, and accelerate toward the elimination of schistosomiasis. *PLoS Neglected Tropical Diseases* 2018 Aug; 12:e0006563. doi: 10.1371/journal.pntd.0006563
59. Amoah AS, Hoekstra PT, Casacuberta-Partal M, Coffeng LE, Corstjens PLAM, Greco B, Lieshout L van, Lim MD, Markwalter CF, Odiere MR, Reinhard-Rupp J, Roestenberg M, Stothard R, Tchuem Tchuente LA, Vlas SJ de, and Dam GJ van. Sensitive diagnostic tools and targeted drug administration strategies are needed to eliminate schistosomiasis. *The Lancet Infectious Diseases* 2020 Jul; 20:e165–e172. doi: 10.1016/S1473-3099(20)30254-1

60. Mazigo HD, Zinga MM, Kepha S, Yard E, McRee-Mckee K, Kabona G, Ngoma DD, and Nshala A. Precision and geographical prevalence mapping of schistosomiasis and soil-transmitted helminthiasis among school-aged children in selected districts of north-western Tanzania. *Parasites & Vectors* 2022 Dec; 15. doi: 10.1186/s13071-022-05547-6
61. Operational Research on Neglected Tropical Diseases (COR-NTD) C for. Schistosomiasis Oversampling Study (SOS): Technical Results Meeting. https://www.eliminatestschisto.org/sites/gsa/files/content/attachments/2023-12-11/SOS%20Nairobi%20Meeting%20Report_Final_eng.pdf. [Accessed on: 2025 Jan 15]
62. World Health Organization. Field use of molluscicides in schistosomiasis control programmes: an operational manual for programme managers. Geneva: World Health Organization, 2017
63. Ciddio M, Mari L, Sokolow SH, De Leo GA, Casagrandi R, and Gatto M. The spatial spread of schistosomiasis: A multidimensional network model applied to Saint-Louis region, Senegal. *Advances in Water Resources* 2017 Oct; 108:406–15. doi: 10.1016/j.advwatres.2016.10.012
64. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, and Buckee CO. Quantifying the Impact of Human Mobility on Malaria. *Science* 2012 Oct; 338:267–70. doi: 10.1126/science.1223467
65. Madsen H and Stauffer JR. Density of *Trematocranus placodon* (Pisces: Cichlidae): A Predictor of Density of the Schistosome Intermediate Host, *Bulinus nyassanus* (Gastropoda: Planorbidae), in Lake Malawi. *EcoHealth* 2011 Jun; 8:177–89. doi: 10.1007/s10393-011-0737-3

4.5 Supplementary Information

Supplementary Tables

Table S4.1. Number of the GPS points per participant per day by district (median, 25th, 75th percentile).

Tercile	Median	Q25	Q75
Mayuge			
bottom	135	87	145
middle	186	171	192
top	226	211	241
Buliisa			
bottom	142	106	170
middle	221	204	237
top	292	269	315
Pakwach			
bottom	236	210	247
middle	277	269	281
top	310	296	326

Table S4.2. Comparison of participants by number of recorded GPS points, grouped into terciles.

Variable	GPS data points per day (tercile)			P-value
	Bottom (n= 151)	Middle (n= 151)	Top (n= 150)	
Gender = male (%)	70 (46.4)	76 (50.3)	80 (53.3)	0.48
Age ≥ 15 years (%)	77 (51.0)	80 (53.0)	97 (64.7)	0.04
Occupation (%)				0.29
none	99 (65.6)	84 (55.6)	76 (50.7)	
fishing	8 (5.3)	17 (11.3)	16 (10.7)	
fishmongering	6 (4.0)	9 (6.0)	7 (4.7)	
farming	28 (18.5)	29 (19.2)	36 (24.0)	
other	10 (6.6)	12 (7.9)	15 (10.0)	
Household drinking water source = improved (%)	75 (49.7)	76 (50.3)	75 (50.0)	0.99
Distance to the closest site (m) (Mean ± SD)	269.0 ± 247.2	283.7 ± 281.2	283.1 ± 281.8	0.86
Distance to the closest tap/borehole (m) (Mean ± SD)	400.3 ± 431.2	392.6 ± 390.4	419.3 ± 425.0	0.85
Radius of gyration (m) (median [IQR])	361.9 [165.0, 1298.6]	371.3 [149.0, 976.2]	325.2 [128.4, 1003.8]	0.56

Notes: All participants were grouped into district-specific terciles (bottom, middle, top) based on their average number of GPS points per day. For binary variables, number and proportions [N (%)] are shown and Chi-squared tests were used to test for group differences. For all normally distributed continuous variables, mean and standard deviation (mean ± sd) are shown and two sample *t*-tests were performed to test for group differences. For all non-normally distributed variables, median and interquartile range (median [IQR]) are shown and Kruskal–Wallis Rank Sum Tests were performed to test for group differences.

Supplementary Figures



Figure S4.1. GPS logger in waterproof pouch and with attached lanyard.



Figure S4.2. GPS loggers with water damage after fieldwork.

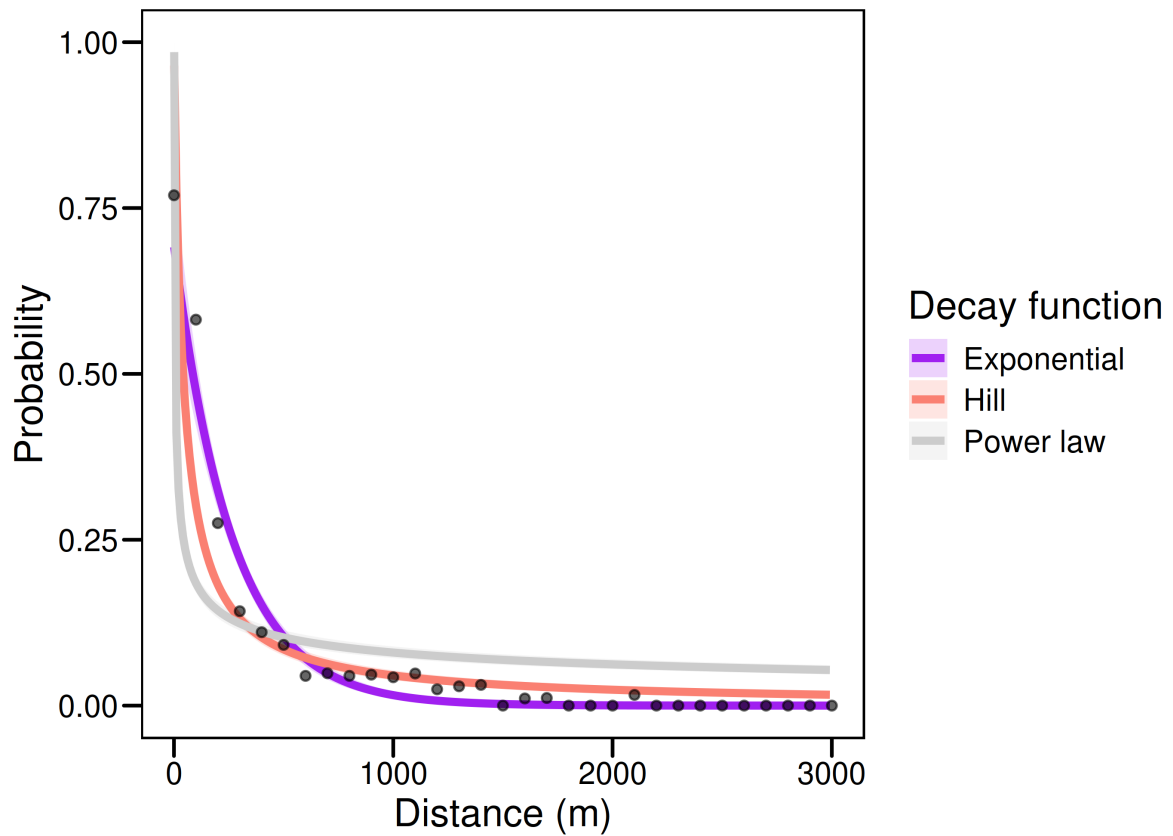


Figure S4.3. Comparison between power law decay, exponential decay, and a Hill function in approximating the global distance decay in water site usage. Black dots represent the true proportion of individuals using water sites, calculated in 100m increments.

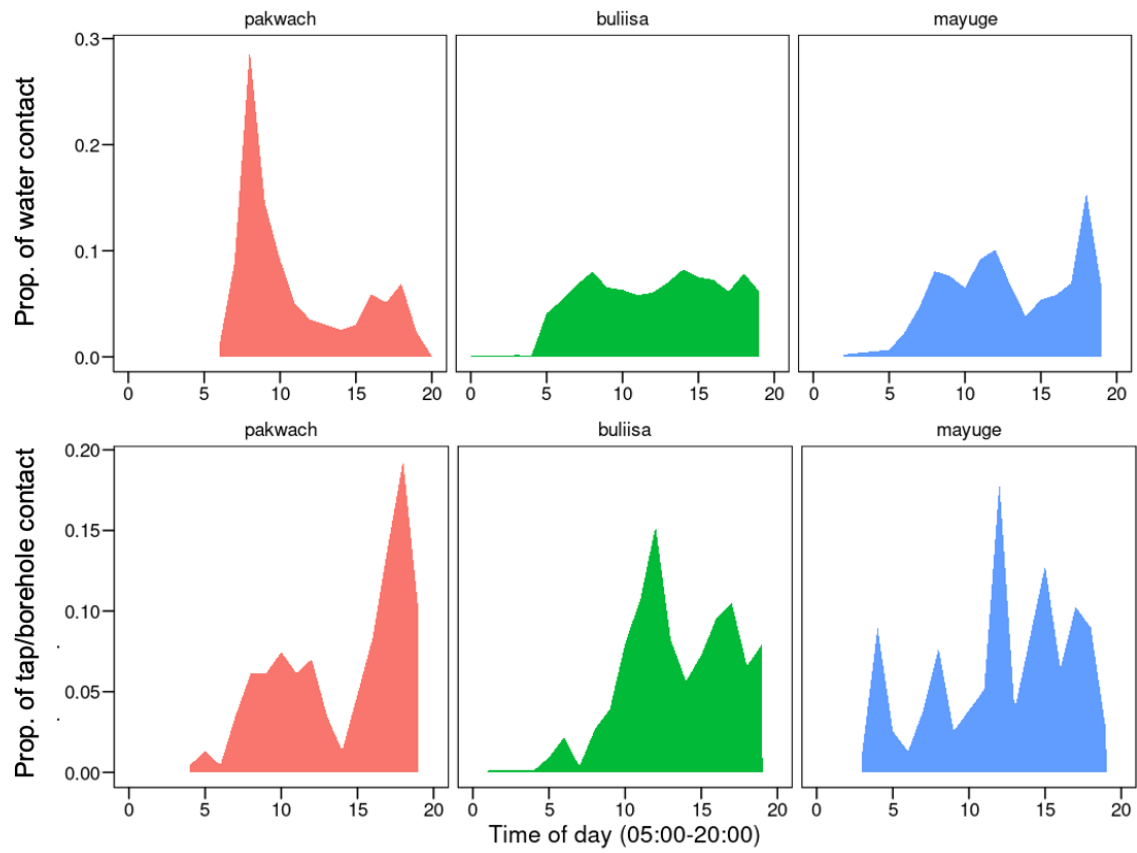


Figure S4.4. Proportion of water site contact and tap/borehole usage episodes by time of day and district.

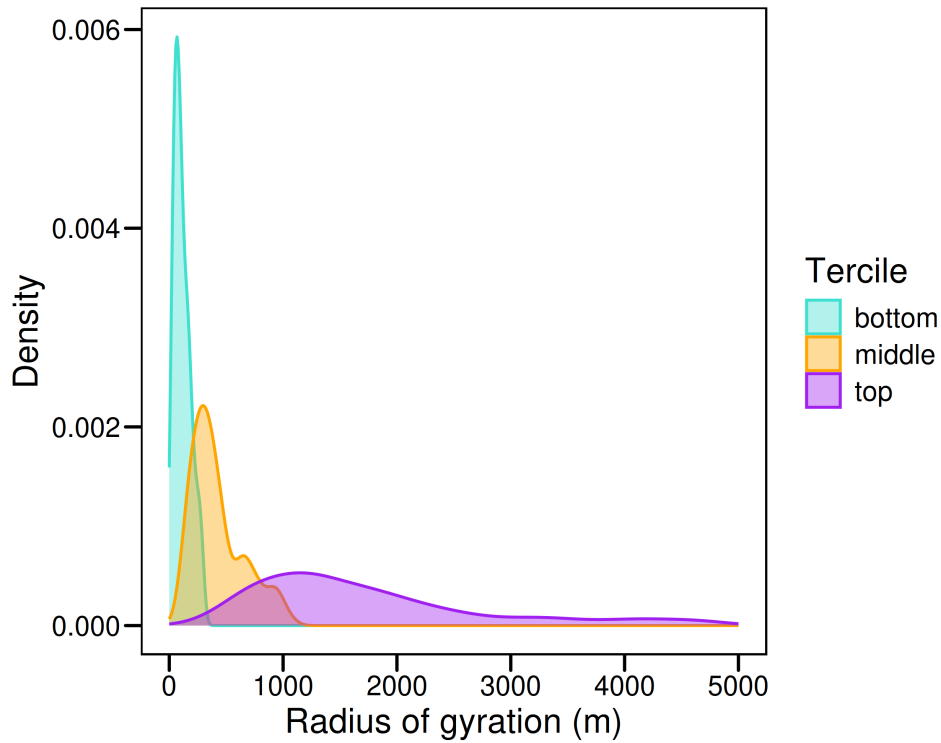


Figure S4.5. Density plot of the district-specific mobility terciles.

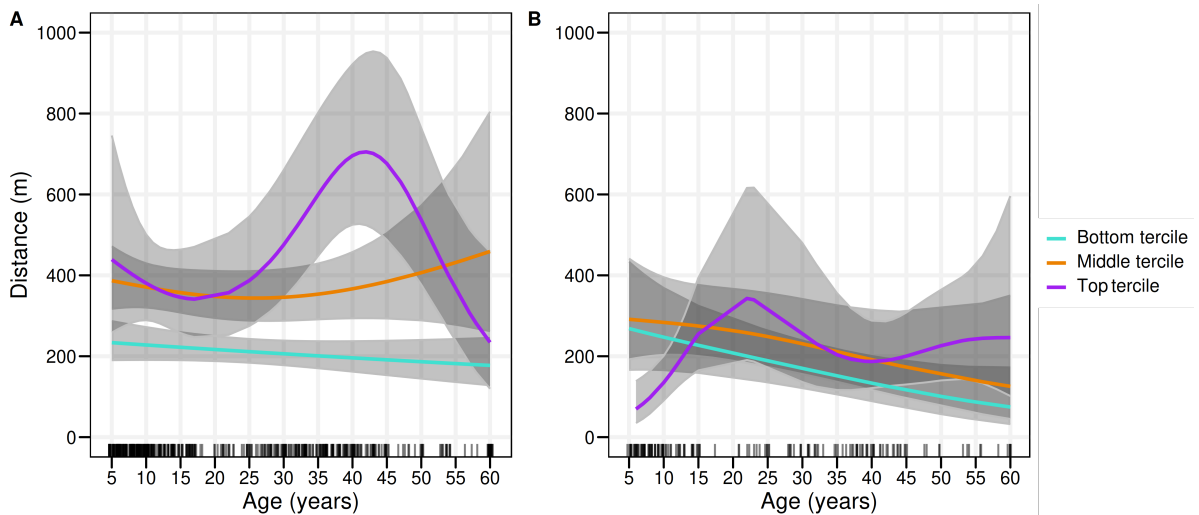


Figure S4.6. GAMs of distance to frequented water sites and taps/boreholes over age and district-specific mobility tercile. Given that only 18 GPS logger participants were > 60 years old, which was too small a sample for modelling mobility-tercile-specific trends over age, we restricted the sample to participants aged ≤ 60 years ($n = 434$) in this figure.

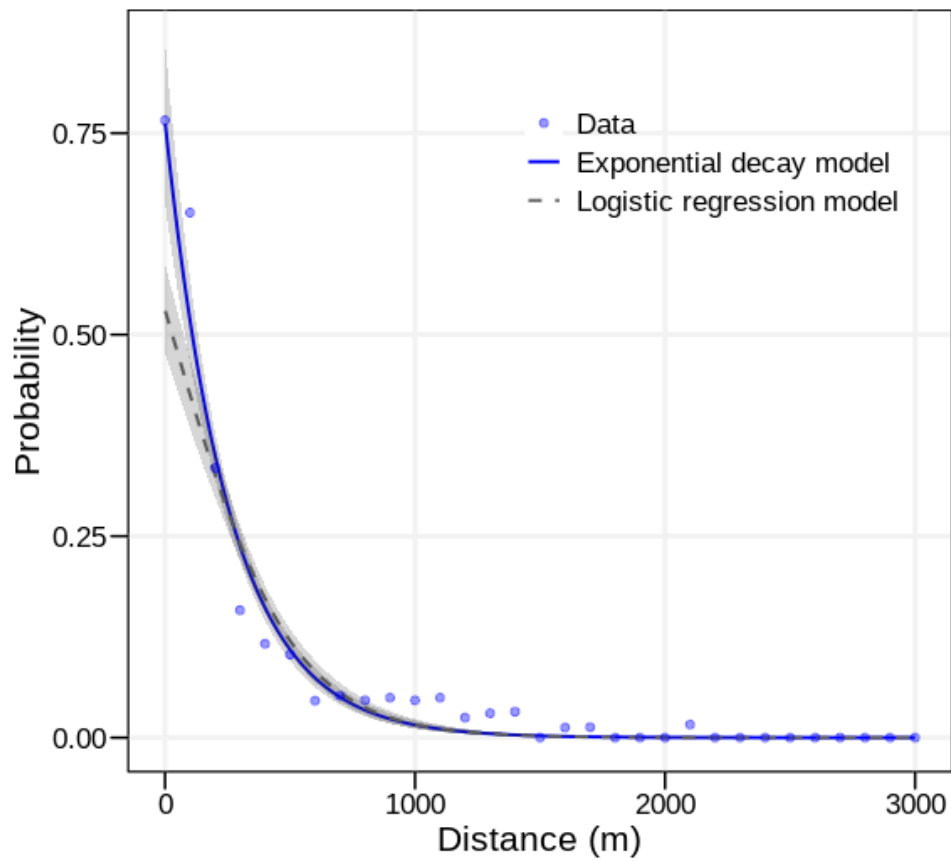


Figure S4.7. Comparison of spatial decay estimated via a logistic regression model versus a spatial decay model with exponential decay function.

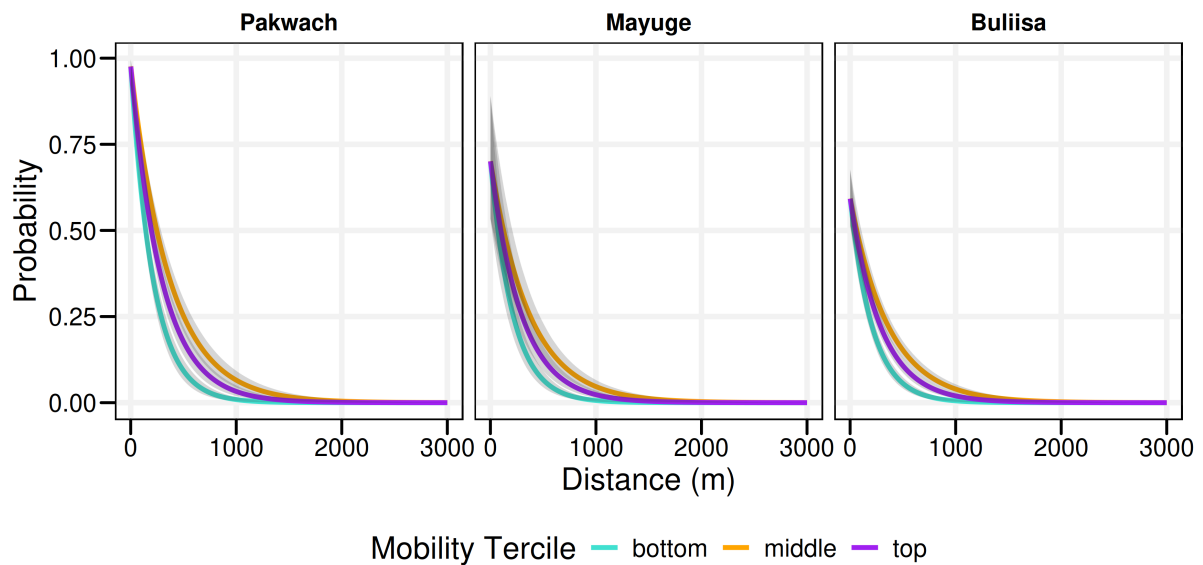


Figure S4.8. Spatial decay models of water site usage incorporating a district-specific intercept and mobility-tercile specific exponential decay.

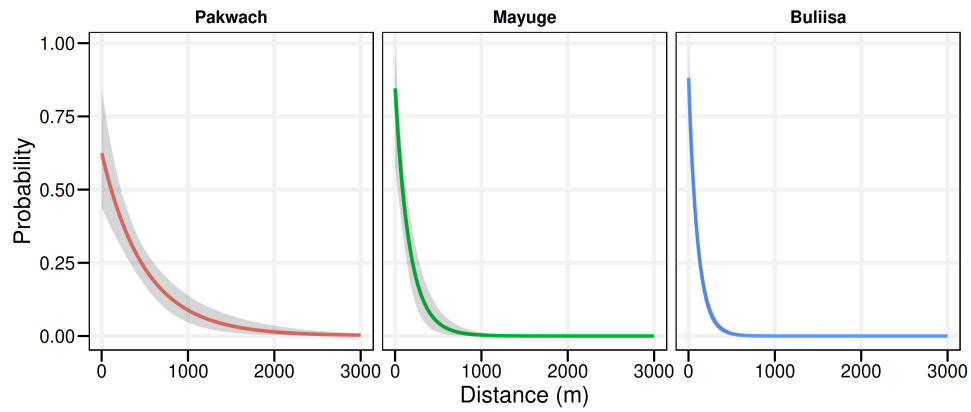


Figure S4.9. Spatial decay models of tap/borehole usage incorporating a district-specific intercept and exponential decay.

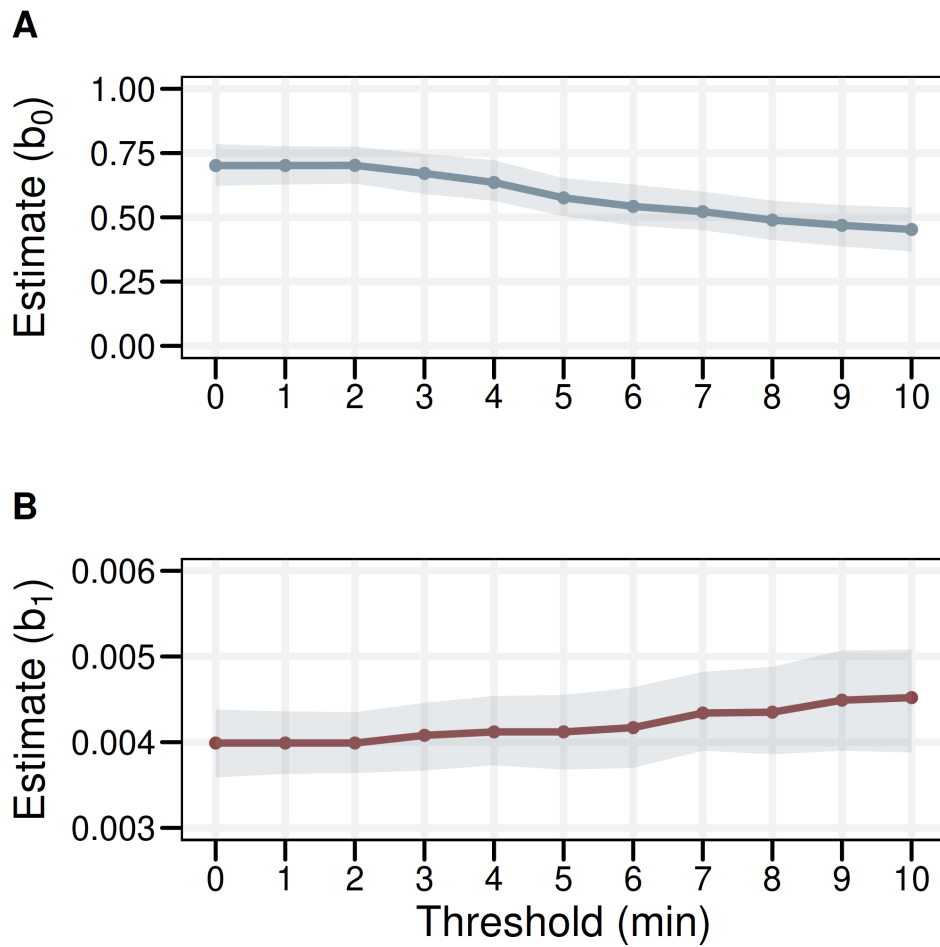


Figure S4.10. The impact of removing short-duration water site and tap/borehole usage events. The threshold value indicates the cut-off value below which events were removed.

Chapter 5

Conclusion

5.1 Thesis Aims and Contributions

5.1.1 Aims

This thesis has furthered our understanding of the roles of current water contact, safe water supplies, and human mobility for the epidemiology of *S. mansoni*, using data from the baseline of the SchistoTrack cohort in rural Uganda. The key questions answered here were: (1) What are the associations between current water contact with *Schistosoma* infection in the published literature? (2) What are the associations between current water contact and infection in SchistoTrack, and are key determinants of water contact and infection different? (3) How do human mobility and safe water supplies influence water site usage behaviours?

5.1.2 Main Contributions

The significance of this work lies in enhancing our understanding of fine-scale human-environmental drivers of schistosome transmission. Major findings from each chapter were as follows:

- Having any self-reported current water contact and engaging in domestic, recreational or domestic water contact was associated with higher odds of schistosome infection in a meta-analysis of 101 studies. There was no evidence that higher water contact duration or frequency increased the odds of infection compared to lower duration or frequency contact (Chapter 2). However, as no standardised exposure measurement tools exist and studies

mostly provided unadjusted effect sizes, reported associations are at risk of confounding.

- In the SchistoTrack study, water contact patterns were highly age- and gender-dependent, and the correlation between current water contact and infection was low. Key determinants of current water contact and infection were largely distinct (Chapter 3).
- Spatial decay models using GPS logger data from SchistoTrack were able to predict individual-level water site and tap/borehole usage. Human mobility influenced water site usage patterns. It was estimated that tap/borehole usage could reduce the probability of water site usage by 10% and water site usage duration by up to 31% (Chapter 4).

These results can inform mathematical modelling and measurement of human behaviour in schistosomiasis research, and guide monitoring and control efforts.

5.2 Summary of Main Findings

5.2.1 The Complex Relationship Between Water Contact and Schistosome Infection

5.2.1.1 *Associations between current water contact and infection were weak*

In SchistoTrack, the association between self-reported current water contact and infection status was very weak. Even before adjustment for any covariates, there was almost no association ($\rho_s = 0.03$). This contrasted with the systematic review meta-analysis in Chapter 2 which found that having any water contact was associated with 3.14 times higher odds of being infected. However, the quality of included studies was largely poor as determined via a critical appraisal, sample sizes were mostly modest and participants were not randomly sampled, and studies did not use validated exposure measurement tools. There was evidence for setting-specific effects, as a positive association was found only in villages with $\geq 10\%$ infection prevalence and in settings with dry-season climates. SchistoTrack provided evidence that there was no particularly strong influence of current water contact on schistosome infection status in rural villages. This was based on multiple pieces of evidence: the raw correlations between water contact and infection were very low, age groups with high infection prevalence had comparatively low levels of water contact and using Bayesian variable selection, it was found that having any

current water contact was not selected as a predictor variable for infection status. Models which allowed for the selection of a rich set of water contact variables to predict infection did not outperform models without any information on individuals' water contact activities. This analysis also included occupation, such as fishing, which is typically indicative of very high levels of exposure. In SchistoTrack, such occupational water contact was associated with both having current water contact and infection status, whereas reporting any water contact was not, illustrating that sustained, intense levels of exposure due to occupation cannot be equated with other water contact activities which are less frequent and shorter duration.

5.2.1.2 *Water contact and schistosomiasis hotspots*

To guide WHO policy, there has been an interest in understanding where and why transmission hotspots exist [1]. The Schistosomiasis Consortium for Operational Research and Evaluation (SCORE) has evaluated the effects of long-term effects of repeated MDA [2]. In SCORE studies, higher levels of water contact have been suggested as one explanation for the persistence of hotspots despite repeated MDA [3, 4]. In the meta-analysis as part of this thesis, associations between water contact and infection increased with community-level prevalence and only existed in settings with dry-season climates. This may point to the relevance of environmental factors as intervening factors in the relationship between water contact and infection. In villages with higher infection prevalence and during the dry season, the likelihood of becoming infected due to a given water contact may be higher if *Schistosoma* parasite presence is higher in these settings [5, 6]. However, studies linking snail variables to human infection outcomes have found mixed results, pointing to the complex human-environmental interactions involved [7–9]. This thesis adds nuance to the role of water contact in hotspot settings by differentiating between the likelihood of individuals having water in a hotspot versus the likelihood of individuals' water contact resulting in infection. For the latter, the meta-analysis results suggest that the likelihood of a given water contact resulting in infection is elevated in higher prevalence settings. At the same time, there was no evidence in SchistoTrack that levels of water contact were higher in villages with higher infection prevalence ($\geq 50\%$) compared to moderate prevalence villages (10–49% prevalence, all by Kato–Katz). While other studies in Côte d'Ivoire and Kenya reported that water contact was higher in persistent hotspot villages [3, 4], multivariable

regression models in Chapter 3 found no evidence for such differences based on village-level prevalence in SchistoTrack.

5.2.2 Key Determinants of Water Contact

5.2.2.1 *Age-specific water contact patterns*

Children are a designated at-risk group in the latest WHO schistosomiasis control guidelines and are assumed to have the highest water contact levels in mathematical transmission models [10–12]. By contrast, in the SchistoTrack study, the proportion of individuals aged < 15 years reporting any water contact was low. We also found that recreational water contact, which has previously been called an important source of exposure [13], was low. In SchistoTrack, water contact patterns of all household members were reported by the household head. It remains possible that recreational water contact was under-reported as such activities are mostly done by children without the presence of adults. At the same time, we triangulated survey and direct observation data which also captures recreational water contact and found that the representation of age groups was similar across both data sources, indicating no evidence for systematic underrepresentation of the water contact behaviours of children.

5.2.2.2 *Gender-specific water contact patterns*

For individuals aged < 15 years, water contact patterns between males and females were similar, suggesting that gender was a relevant determinant of water contact patterns only in adults. Among adults (aged ≥ 15 years), there were clear gender differences in the composition of their water contact activities. Female adults predominantly engaged in domestic water contact, whereas males were mostly involved in occupational water contact. In multivariable statistical models, we found that the gender effect varied based on which type of water contact was predicted. Females had significantly higher odds of domestic water contact compared to males while having significantly lower odds of occupational and recreational water contact. These gendered activity patterns did translate into differences in different intensities of water contact. Females had a higher number of weekly trips to water sites, as would be expected with activities such as getting drinking water or washing laundry, while males had a higher overall duration as they engaged in long-duration activities such as fishing.

5.2.2.3 *Impact of human mobility on water contact*

Using individual-level GPS logger data, there was a non-linear relationship between water contact and human mobility. For people in the lowest mobility tercile, distances to frequented water sites tended to be consistently lower compared to individuals in the middle and top terciles, indicating that low mobility range may have constrained which water sites they visited. However, the relationship was non-monotonic as people in the highest mobility tercile did not visit sites further away than people in the middle tercile. Across age, the highest mobility was observed for individuals between 20–40 years, which coincided with the period of highest water contact. This thesis provides possibly the first evidence for the relevance of fine-scale individual-level mobility for water contact patterns. The findings do not support the idea that highly mobile individuals necessarily have more water contact. This finding aligns with studies on larger-scale mobility patterns which suggested that higher mobility is not necessarily related to higher infection risk or transmission [14, 15]. Nonetheless, mobile individuals likely play a key role in long-distance transmission, whereas low-mobility individuals may have their water contact sites significantly constrained by their limited movement range.

5.2.3 Environmental and Spatial Factors

5.2.3.1 *Distance to waterbodies*

Distance to waterbodies was found to be a significant factor influencing water contact patterns, as demonstrated by both survey and GPS data. The self-reported data where individuals were assigned to the closest site revealed a linear decline in self-reported water contact. The analysis using the more granular GPS data, where contacts with all local, field-mapped water sites were recorded, showed an exponential decay in water contact over distance. The assumption frequently made in survey data that individuals only use the closest site is mostly warranted as the GPS data revealed that 85% of the total site usage duration was with the closest site. There was strong evidence that water site usage decayed rather quickly over distance; in survey data 80% of all individuals with water contact were found to live within 430m of waterbodies, while in GPS data, 84% of used sites were within 500m of individuals' households. Compared to water site usage, tap/borehole usage was even more focal with 89% of individuals using taps/boreholes

within 500m of the household. These findings confirm the strong spatial dependence of water contact and tap/borehole usage on small spatial scales.

5.2.3.2 *Water site characteristics and contamination*

There was some evidence that characteristics of water sites influenced water contact patterns. In SchistoTrack, people who lived close to beaches, swamps or ponds were more likely to self-report water contact compared to those living closest to a river. Among spatial decay models, whether a self-reported water contact activity matched one or more of the activities typically performed at the site was one of the better-performing models, but was not selected as the best model. There was evidence that local presence of faecal contamination at water sites influenced water contact. Individuals were less likely to self-report water site usage when the closest water site was contaminated. This could reflect true site avoidance behaviour whereby contamination reduces the activities performed at the site. Alternatively, this may also be driven by social desirability bias if individuals are less likely to report visiting sites in areas with substantial contamination. Analyses which used variables for whether snails or infected snails were present at a water site found that these factors did not influence self-reported water contact, as would be expected, given that snail abundance and infection rates are not easily observable by individuals.

5.2.4 WASH Factors

5.2.4.1 *Safe sanitation infrastructure*

There was no strong evidence that access to safe sanitation reduced open water contact within SchistoTrack. Among the two sanitation-related variables included in the Chapter 3 analysis—presence of a home latrine, and household distance to the closest public latrine—only the latter was selected for inclusion in the water contact model. Individuals living closer to public latrines had higher odds of water contact. This association was likely confounded by distance as individuals with high water contact typically lived close to waterbodies as well as public latrines which were mostly built close to lakes. The absence of strong associations between public sanitation and water contact could be driven by comparatively high population-wide private sanitation coverage across all study areas. Among participants, 85% reported having a home latrine. At the same time, 85% coverage may not be sufficient for private sanitation

to suppress transmission as near-universal coverage and usage may be required for sanitation to reduce water contact and cercarial exposure [13, 16]. (In this thesis, only coverage, but not usage, of sanitation infrastructure was measured.)

5.2.4.2 *Safe water supplies*

From survey-based variables of safe water access and use—household usage of taps/boreholes as the main drinking water source, household purification of drinking water, and distance to the closest tap/borehole—none were selected for inclusion. Using the GPS logger data, we were able to obtain individual-level measures of tap/borehole usage. The descriptive results showed that among people using either water sites, taps/boreholes, or both, there was a negative association between site usage and safe water supply usage. This was affirmed using spatial decay models with crowding, where it was estimated that the presence of taps/boreholes reduced the probability of water site usage by 10% and duration of usage by 31%. The findings of the survey-based analysis versus the GPS logger analysis are not necessarily contradictory as survey-based measures of safe water supplies were collected at the household level and did not reflect individual-level usage behaviours. The magnitude of the crowding-out effect reported here represents an upper bound of the influence of taps/boreholes on open water contact as the model selection results provided no strong evidence for preferring the decay models with crowding out over models without crowding out.

5.3 Implications for Water Contact Measurement, Modelling, and study design

5.3.1 Water Contact Measurement

Currently, there exists no standardised tool for water contact measurement in the context of schistosomiasis [16]. This thesis highlights that appropriate measurement tools and units (any water contact, frequency, duration) are dependent on the specific research question or implementation context.

5.3.1.1 *Any water contact*

Chapter 3 demonstrated that self-reported measures of any water contact, collected by asking respondents for their typical activities, can yield measures which adequately capture community-level water contact patterns, as shown via triangulation with direct observation data. Moreover, broad umbrella categories, such as grouping water contact into recreational, occupational, and domestic water contact, were useful for capturing age- and gender-related trends in water contact. As age, gender, and occupation were significant predictors of water contact, collecting basic demographic data may be a feasible way of obtaining proxy measures of having any water contact in control programmes, or for designing exposure reduction interventions that target specific at-risk groups for water contact. At the same time, as we show in Chapter 3, measures of current water contact for individuals are not necessarily indicative of their infection outcomes.

5.3.1.2 *Water contact frequency and duration*

Frequency and duration measures are important for understanding dose-response relationships between exposure and infection [17, 18]. In the systematic review in Chapter 2, there was no clear evidence that higher intensity of water contact was linked to higher odds of infection, based mostly on self-reported data. Statistical heterogeneity, as measured by the I^2 statistic, increased with higher frequencies, indicating a potential challenge in adequately capturing more frequent water contact across different studies with different measurement tools. Within the SchistoTrack data, self-reported measures provided internally consistent information on frequency and duration. Fishing, for instance, had the highest average duration while bathing and getting drinking water had the shortest durations, which aligns with what has been found elsewhere using direct observation [19]. However, as it is difficult for participants to accurately estimate frequencies and durations, these measures are likely to be somewhat noisy. Direct water contact observation data and GPS-derived contact promise to provide more reliable and consistent estimates. Observation data has been mostly used to obtain aggregate-level measures of frequency and duration because it is challenging to derive individual-level metrics from such site-level data as this requires identifying individuals across data entries, usually hand-written by different observers without homogeneous spelling of names. This thesis demonstrates that by combining GPS data with water site location data, individual-level measures of water contact

duration can be derived. Importantly, this data can capture short contact events, which is important given that in observation studies, daily durations of water contact in endemic settings were typically low (between 4–12 min per day) [19, 20]. Obtaining frequency estimates from GPS data at water sites is more challenging compared to duration because the GPS locations of an individual may oscillate between being inside and outside the buffer area of a site when a small buffer size is chosen. It thus remains difficult to identify what constitutes a single versus multiple separate contact events, but reliable classification of distinct events is needed to estimate frequencies. It should also be noted that measures of duration or frequency may capture slightly different aspects of water contact across measurement tools. In observation data, a water contact is usually defined as being at the water site, even when individuals are not immersed in water, whereas in GPS data, individuals need to be inside a narrowly defined buffer, which means that the duration of water contact will typically be shorter in GPS data. Due to these differences, one may not expect to have perfect agreement between durations in direct observation and GPS data. Even for duration and frequency, there is not necessarily one superior measurement tool; surveys, direct observation, and GPS loggers all may be useful, depending on the research question.

5.3.2 Modelling Water Contact

5.3.2.1 Predictability and determinants of water contact

The results from Chapter 3 show that having any water contact can be successfully predicted using socio-demographic and environmental variables. Despite the complexity of human behaviour, models predicting any water contact had a high model performance (auROC = 0.777), which exceeded that of models predicting Schistosome infection status (auROC = 0.693). The interrelatedness of determinants of water contact and infection calls for more careful modelling approaches that try to disentangle the pathways through which risk factors operate. For instance, in the case of gender, it has been found that males were at higher risk than females for infection, but it is unclear whether this is due to differences in susceptibility, different exposure patterns, or other factors [18, 21]. In SchistoTrack, females were more likely to have water contact, but only after occupation was accounted for. At the same time, gender was not selected as a predictor

for the infection status model. Based on these results, it seems likely that gender differences in infection may be explained by differences in water contact behaviour as opposed to biological differences in susceptibility. More work, especially using directed acyclic graphs, is needed to clarify the roles and relationships between risk factors of water contact and infection.

5.3.2.2 *Using water site distance for predicting water contact*

Due to the strong household distance dependence of water contact behaviour, spatial models provide a powerful tool for modelling water contact behaviour using simple mathematical tools. As shown in Chapter 3, distance measures were not interchangeable as water contact decayed over household water site distance but not school distance to water sites. Village centre distance was found to truncate waterbody distances of the households furthest away from sites. In spatial decay models, the usage of water sites and taps/boreholes and their respective usage duration could be predicted based on household distance. Spatial decay models require only low numbers of GPS traces to estimate the decay parameters and predict population-wide site usage patterns.

5.4 Implications for Control Strategies

The findings from this thesis have implications for major control strategies: MDA, exposure reduction interventions, environmental control, and safe water provision.

5.4.1 MDA

Due to the weak association between water contact and infection in SchistoTrack, having any water contact cannot serve as a simple proxy measure for identifying infected individuals across different contexts. While the meta-analysis showed a significant relationship between water contact and infection, the SchistoTrack data as well as another study from Uganda found no significant association [22]. This heterogeneity in exposure-infection associations support continued use of community-wide mass drug administration in highly endemic areas regardless of individuals' water contact patterns. This thesis also provides a justification for including adults in MDA as they had the highest levels of water contact and are thus at risk of being (re-)infected. The SchistoTrack study focused on households very close to the shoreline ($< 2.2\text{km}$) and thus investigated trends in water contact on very small spatial scales. This is smaller than scales

typically used in operational research and implementation. For instance, the Schistosomiasis Oversampling Study led by the Global Schistosomiasis Alliance focused on < 5km waterbody distance as the relevant distance metric [23]. As we focused on highly local scales, it remains possible that clear relationships between water contact and infection exist at larger spatial scales and that individuals without water contact could be treated less frequently to reduce overtreatment while focusing medicines on the most at-risk segments in endemic areas. This is particularly important given current funding shortfalls for continued MDA in Uganda and elsewhere [24].

5.4.2 Exposure Reduction Interventions

In the past, behaviour change strategies have focused on children due to their high levels of infection [25, 26]. The high levels of water contact in adults demonstrated in SchistoTrack justify programmes that focus on reducing exposure across all age groups. However, as different groups had different water contact patterns—males mostly occupational water contact and females and children domestic water contact—exposure reduction interventions are unlikely to succeed unless they target specific behaviours of these groups. For instance, safe water infrastructure will likely reduce the water contact of female adults and children, but will likely not impact water contact of male adults. Equitable interventions that target major sources of water contact for all different demographic groups are required to achieve community-wide reductions in exposure.

5.4.3 Environmental Control

Spatial decay models provide novel opportunities for more effectively implementing mollusciciding or other environmental control approaches. With comparatively little GPS data and limited field mapping of households and water sites, these models can predict population-level water contact patterns and help to identify priority water sites for control efforts. The WHO guidelines state that when using focal control, all water sites used by the community should be treated, assuming the individuals use mostly sites within their village [27]. GPS data provides human mobility information and thus yields information on how far individuals travel to visit

sites, providing ground-truth information on the radius in which sites need to be treated to control environmental sources of exposure for a given community.

5.4.4 Safe Water Supplies

The spatial decay models with crowding out presented in this thesis can provide estimates of the impact of safe water supplies on community water contact. The results show that taps/boreholes are more likely reducing, rather than completely eliminating, open water contact. However, as shown in the meta-analysis, even relatively low levels of water contact (having any water contact), may be sufficient to become infected especially for children, illustrating the challenge for WASH to sufficiently suppress open water contact. Chapter 4 also demonstrated how such spatial decay models can predict the impact of potential interventions that increase access to safe water supplies. This exercise revealed that in the SchistoTrack study communities, individuals living close to water sites typically had high levels of access to safe public water supplies, meaning that there was limited potential of building public infrastructure to further reduce exposure. Even under scenarios with drastic increases in tap/borehole access, community-level water contact durations did not approach zero. Thus, public water supplies may not be sufficient for controlling schistosomiasis. Private water supplies have shown more promise. A study from St Lucia provided evidence for the benefits of private water supplies as the installation of household water supplies significantly reduced schistosome prevalence, incidence, and intensity whereas prevalence increased in comparison areas where only public standpipes were provided [28].

5.5 Generalisability

The focus of this thesis was on understanding water contact patterns in sub-Saharan Africa. In the systematic review, 77% of all included studies were from the African continent, while the SchistoTrack study focused exclusively on three Ugandan districts. While the meta-analysis found an association between water contact and infection, there was no such relationship in SchistoTrack. This raises questions about whether the findings from SchistoTrack generalise to other geographical settings in Uganda and beyond. SchistoTrack found no significant impact of socio-demographic factors, such as religion or tribe, on water contact. However, these factors

are region-specific, and further research is needed to assess their role in other contexts. As Uganda is endemic to *S. mansoni*, this thesis focused on water contact patterns as a determinant of *S. mansoni* infection. However, especially for *S. haematobium*, where environmental contamination occurs via open urination, the key risk groups and behaviours are likely different, as urination by children, and boys in particular, may be a key contributor to *S. haematobium* transmission [29]. Immune responses may also differ between *S. mansoni* and *S. haematobium* [30], calling for longitudinal studies on exposure and infection for both *S. mansoni* and *S. haematobium* to draw general conclusions about the role of water contact in schistosome infection.

5.6 Thesis Limitations

This thesis has several limitations. In the systematic review and meta-analysis in Chapter 2, statistical heterogeneity was high across studies, reflecting both heterogeneous measurement and study settings included. It was difficult to ascertain the reliability of water contact measurement across studies, especially as tools and recall periods varied widely. The literature searches in Chapter 2 were limited to peer-reviewed studies which excluded the grey literature. Studies were also limited to the English language, excluding studies published in other languages, especially Portuguese and French.

SchistoTrack data showed that even within the same participants, agreement between GPS-derived and self-reported water contact was low. Different measurement tools can capture different facets of water contact behaviour. Surveys best measure typical water contact patterns; direct observation or GPS loggers measure water contact in a specific time period that may not always be representative of an individual's general water contact patterns.

This thesis focused on the role of water supplies for water contact behaviour. While a few sanitation variables were included in Chapter 3, a detailed investigation of the influence of sanitation and open defecation behaviour on water contact was beyond the scope of the thesis.

This thesis focused on providing a quantitative understanding of the roles of human water contact and behaviour in schistosome transmission. However, qualitative data collection is needed to better characterise the social and cultural factors that shape water contact, as quan-

titative data alone is insufficient to provide a rich understanding of the local context. On Lake Albert, qualitative research demonstrated how common beliefs, such as lake water having a better taste or faeces acting as bait for fish, shape the WASH behaviour people engage in [31]. Consequently, future studies should deploy qualitative research alongside quantitative data collection to understand the drivers of behaviour and design effective control interventions that address people’s needs and concerns.

SchistoTrack data captured water contact patterns and mobility across three diverse geographic districts. However, as water contact-infection associations differed from the meta-analysis findings, it is unclear to which extent the findings apply to other regions and countries. The findings here are in line with a study by Exum et al. in Uganda [22] yet whether this indicates country- or region-specific water contact patterns requires further investigation.

The SchistoTrack study was extremely fine-scale and thus focused on understanding water contact and infection gradients at very local scales in high endemicity settings. Studies at larger geographic scales are required to investigate gradients at larger scales. Other research, such as the Schistosomiasis Oversampling Study, considered 5 km relevant for water contact and Exum et al. included households more than 10 km away from waterbodies [22, 23]. While we found no obvious gradients of infection over waterbody distance, it is highly likely that such gradients exist at larger scales.

While this study used Bayesian variable selection to identify the key human-environmental drivers of water contact, these factors may be specific to our study setting. For instance, infection prevalence has been shown to be higher for males than females in some studies but not others [21], potentially driven by different gendered water contact patterns across settings. This study included different religious and tribal groups and found no significant impact of these factors on water contact, but more research is needed to understand how setting-specific the impacts of—for instance—demographic factors are.

SchistoTrack included detailed human-environmental data but lacked infection history and immunological data. Therefore, this thesis was unable to account for individuals’ immune responses, the timing of first infection, longevity of infections, co-infections, and detailed immune responses. There are outstanding research questions around the build-up of immunity

due to past infection, the relevance of cumulative exposure, and changes in individuals' water contact over time, for instance, measured via repeated water contact surveys, that are beyond the scope of the available SchistoTrack data.

5.7 Conclusion and Future Directions

This thesis focused on analysing the roles of current water contact, safe water supplies, and human mobility for the epidemiology of *S. mansoni* in rural Uganda by synthesising the existing literature and using original field data from the SchistoTrack study. The results highlight the complex relationship between water contact and infection, identified the distinct determinants of water contact and infection, and provided evidence for the relevance of human mobility and the role of safe water supplies for water contact behaviour. As the drivers of schistosomiasis are multifaceted—encompassing human-behavioural, environmental, socio-demographic, and biological factors—this thesis provides a template for how interdisciplinary research drawing upon field data can interrogate the relevance of these factors in the epidemiology of schistosomiasis.

The results provide a novel set of assumptions about the roles of water contact, human behaviour, and safe water supplies for schistosomiasis. (1) Having current water contact is not necessarily a strong predictor of infection outcomes. (2) Age-specific water contact patterns do not approximate age-specific infection rates. (3) Immunity is likely very important for determining age-dependent infection prevalence. (4) Gender differences in water contact are largely due to gender differences in occupational and domestic water contact patterns. (5) Water contact is not linearly related to mobility. (6) Site usage behaviours can be approximated using exponential decay functions. (7) Public water supplies can reduce duration of open water contact but are unlikely to completely crowd out open water contact.

Findings from this thesis open up multiple avenues for future research. Here the focus was demonstrating that GPS loggers can estimate and predict small-scale patterns of water contact. GPS data from loggers may potentially also capture larger-scale movements. This is important because mobile and migrant populations are often not reached by MDA programmes but can play important roles in transmission, especially in near-elimination settings [32, 33]. Environmental changes and human displacement due to climate change are another emerging field of research

where GPS movement data, combined with high resolution-remote sensing data, could enable research to understand the cascading effects of climate change on schistosome transmission across spatial scales. However, a key challenge for this research is on how to upscale GPS studies. The limited battery life, high cost, and fieldwork needed to implement such studies are currently a barrier for larger-scale and longer-duration studies. Therefore, experimentation with alternative technologies, such as smartphones—which are ubiquitous even in low-income countries—will be needed.

This thesis provides evidence that current water contact and infection status are largely unrelated. There is a need for longitudinal exposure-infection studies which incorporate multiple exposure assessment points to understand persistence and variability in water contact over time. Ideally, such studies would combine exposure histories with immunological assessments to confirm the relative importance of acquired immunity versus exposure for infection dynamics, as this thesis suggests a potentially important role of acquired immunity for determining infection outcomes over the life course.

This thesis enables the development of mathematical schistosome transmission models with more realistic assumptions about human behaviour. While current models such as SCHISTOX or SchiSTOP assume that exposure peaks in school-age children [12, 34], this thesis suggests the need for models that incorporate a peak of exposure in adults and possibly stronger within-host regulating mechanisms, such as acquired immunity. The crowding out analysis for water site usage supports the future development of individual-based models as an adequate framework for studying water contact and infrastructure development, and if extended to re-infection, then also more granular schistosome transmission modelling. There also is a need for meta-population models that draw upon the empirical patterns of water site usage and human mobility observed here to better understand where individuals are getting infected and, if possible, to assign individuals to the site of infection acquisition. It should be investigated whether such meta-population models, when incorporating human behavioural data and site-level variation in environmental risk, provide more adequate accounts of the small-scale heterogeneities in village-level prevalence observed in studies like SchistoTrack and elsewhere [35–37].

To summarise, the public health implications of this thesis lies in characterising risk be-

haviours and identifying potential strategies to reduce open water contact. Such information is needed to complement MDA with additional control strategies including WASH and behaviour change which are needed as MDA alone is unlikely to achieve elimination as a public health problem [34]. In SchistoTrack, adults had some of the highest levels of water contact, highlighting the importance of population-wide MDA and exposure reduction interventions. As water contact of male adults was mostly occupational, multifaceted strategies that target both domestic activities of women and children and occupational fishing and agriculture of men are needed. Spatial proximity was a strong driver for both open and safe water contact; therefore interventions which increase the relative accessibility of public safe water supplies compared to water sites may reduce exposure. At the same time, safe water supplies are unlikely to completely suppress open water contact. Understanding water contact patterns can also help to divert contacts to times outside of peak cercarial shedding hours, treat commonly used sites with molluscicides, and identify optimal locations for building public water infrastructure. This thesis highlights the importance of detailed behavioural data for better understanding the epidemiology and non-pharmaceutical approaches to schistosomiasis control.

References

1. Wiegand RE, Mwinzi PNM, Montgomery SP, Chan YL, Andiego K, Omedo M, Muchiri G, Ogutu MO, Rawago F, Odiere MR, Karanja DMS, and Secor WE. A Persistent Hotspot of *Schistosoma mansoni* Infection in a Five-Year Randomized Trial of Praziquantel Preventative Chemotherapy Strategies. *The Journal of Infectious Diseases* 2017 Sep; 216:1425–33. doi: 10.1093/infdis/jix496
2. University of Georgia. Schistosomiasis Consortium for Operational Research and Evaluation. <https://ctegd.uga.edu/score/>. [Accessed on: 2025 Jan 27]
3. Assaré RK, N'Tamon RN, Bellai LG, Koffi JA, Mathieu TBI, Ouattara M, Hürlimann E, Coulibaly JT, Diabaté S, N'Goran EK, and Utzinger J. Characteristics of persistent hotspots of *Schistosoma mansoni* in western Côte d'Ivoire. *Parasites & Vectors* 2020 Jul; 13. doi: 10.1186/s13071-020-04188-x
4. Musuva R, Odiere M, Mwinzi P, Omondi I, Rawago F, Matendechero S, Kittur N, Campbell C, and Colley D. Unprotected water sources and low latrine coverage are contributing factors to persistent hotspots for schistosomiasis in western Kenya. *PLoS ONE* 2021 Sep; 16. doi: 10.1371/journal.pone.0253115
5. Hailegebriel T, Nibret E, and Munshea A. Distribution and seasonal abundance of *Biomphalaria* snails and their infection status with *Schistosoma mansoni* in and around Lake Tana, northwest Ethiopia. *Scientific Reports* 2022 Oct; 12. doi: 10.1038/s41598-022-21306-0
6. Gouvras AN, Allan F, Kinung'hi S, Rabone M, Emery A, Angelo T, Pennance T, Webster B, Nagai H, and Rollinson D. Longitudinal survey on the distribution of *Biomphalaria sudanica* and *B. choanomophala* in Mwanza region, on the shores of Lake Victoria, Tanzania: implications for schistosomiasis transmission and control. *Parasites & Vectors* 2017 Jun; 10. doi: 10.1186/s13071-017-2252-z
7. Chandiwana SK. Community water-contact patterns and the transmission of *Schistosoma haematobium* in the highveld region of Zimbabwe. *Social Science & Medicine* 1987 Jan; 25:495–505. doi: 10.1016/0277-9536(87)90173-0

CHAPTER 5. CONCLUSION

8. Chandiwana SK, Taylor P, and Clarke VDV. Prevalence and intensity of schistosomiasis in two rural areas in Zimbabwe and their relationship to village location and snail infection rates. *Annals of Tropical Medicine & Parasitology* 1988 Jan; 82:163–73. doi: 10.1080/00034983.1988.11812224
9. Wood CL, Sokolow SH, Jones IJ, Chamberlin AJ, Lafferty KD, Kuris AM, Jocque M, Hopkins S, Adams G, Buck JC, Lund AJ, Garcia-Vedrenne AE, Fiorenza E, Rohr JR, Allan F, Webster B, Rabone M, Webster JP, Bandagny L, Ndione R, Senghor S, Schacht AM, Jouanard N, Riveau G, and De Leo GA. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. *Proceedings of the National Academy of Sciences* 2019 Oct; 116:23182–91. doi: 10.1073/pnas.1903698116
10. World Health Organization. WHO guideline on control and elimination of human schistosomiasis. Geneva: World Health Organization, 2022
11. Anderson R, Turner H, Farrell S, and Truscott J. Studies of the Transmission Dynamics, Mathematical Model Development and the Control of Schistosome Parasites by Mass Drug Administration in Human Communities. *Mathematical Models for Neglected Tropical Diseases - Essential Tools for Control and Elimination, Part B*. Elsevier, 2016 :199–246. doi: 10.1016/bs.apar.2016.06.003
12. Graham M, Ayabina D, Lucas TC, Collyer BS, Medley GF, Hollingsworth TD, and Toor J. SCHISTOX: An individual based model for the epidemiology and control of schistosomiasis. *Infectious Disease Modelling* 2021; 6:438–47. doi: 10.1016/j.idm.2021.01.010
13. Grimes JE, Croll D, Harrison WE, Utzinger J, Freeman MC, and Templeton MR. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasites & Vectors* 2015 Dec; 8:156. doi: 10.1186/s13071-015-0766-9
14. Buchwald AG, Grover E, Van Dyke J, Kechris K, Lu D, Liu Y, Zhong B, and Carlton EJ. Human Mobility Associated With Risk of *Schistosoma japonicum* Infection in Sichuan, China. *American Journal of Epidemiology* 2021 Jan. doi: 10.1093/aje/kwaa292
15. Mari L, Gatto M, Ciddio M, Dia ED, Sokolow SH, De Leo GA, and Casagrandi R. Big-data-driven modeling unveils country-wide drivers of endemic schistosomiasis. *Scientific Reports* 2017 Mar; 7. doi: 10.1038/s41598-017-00493-1
16. Reitzug F, Ledien J, and Chami GF. Associations of water contact frequency, duration, and activities with schistosome infection risk: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2023 Jun; 17:e0011377. doi: 10.1371/journal.pntd.0011377
17. Wilkins H, Blumenthal U, Hagan P, Hayes R, and Tulloch S. Resistance to reinfection after treatment of urinary schistosomiasis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1987 Jan; 81:29–35. doi: 10.1016/0035-9203(87)90273-2
18. Bundy D. Gender-dependent patterns of infections and disease. *Parasitology Today* 1988 Jul; 4:186–9. doi: 10.1016/0169-4758(88)90076-2
19. Sow S, Vlas SJ de, Stelma F, Vereecken K, Gryseels B, and Polman K. The contribution of water contact behavior to the high *Schistosoma mansoni* Infection rates observed in the Senegal River Basin. *BMC Infectious Diseases* 2011 Dec; 11:198. doi: 10.1186/1471-2334-11-198
20. Kloos H, Rodrigues JCAP, Pereira WR, Velásquez-Meléndez G, LoVerde P, Oliveira RC, and Gazzinelli A. Combined methods for the study of water contact behavior in a rural schistosomiasis-endemic area in Brazil. *Acta Tropica* 2006 Jan; 97:31–41. doi: 10.1016/j.actatropica.2005.08.006
21. Ayabina DV, Clark J, Bayley H, Lamberton PHL, Toor J, and Hollingsworth TD. Gender-related differences in prevalence, intensity and associated risk factors of *Schistosoma* infections in Africa: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases* 2021 Nov; 15:e0009083. doi: 10.1371/journal.pntd.0009083
22. Exum NG, Kibira SPS, Ssenyonga R, Nobili J, Shannon AK, Ssempebwa JC, Tukahebwa EM, Radloff S, Schwab KJ, and Makumbi FE. The prevalence of schistosomiasis in Uganda: A nationally representative population estimate to inform control programs and water and sanitation interventions. *PLoS Neglected Tropical Diseases* 2019 Aug; 13:e0007617. doi: 10.1371/journal.pntd.0007617
23. Alliance GS. Schistosomiasis Oversampling Study: Survey Strategy Selection. <https://www.eliminate-schisto.org/sites/gsa/files/content/attachments/>. [Accessed on: 2025 Jan 22]

24. Africa WRO for. Act Now, Save Lives: Ensuring Lifesaving Medicines Reach Uganda's Most Vulnerable. <https://espen.afro.who.int/updates-events/updates/act-now-save-lives-ensuring-lifesaving-medicines-reach-uganda%E2%80%99s-most>. [Accessed on: 2024 Apr 16]
25. Knopp S, Mohammed KA, Ali SM, Khamis IS, Ame SM, Albonico M, Gouvras A, Fenwick A, Savioli L, Colley DG, Utzinger J, Person B, and Rollinson D. Study and implementation of urogenital schistosomiasis elimination in Zanzibar (Unguja and Pemba islands) using an integrated multidisciplinary approach. *BMC Public Health* 2012 Oct; 12. doi: 10.1186/1471-2458-12-930
26. Knopp S, Person B, Ame SM, Ali SM, Hattendorf J, Juma S, Muhsin J, Khamis IS, Mohammed KA, Utzinger J, Hollenberg E, Kabole F, and Rollinson D. Evaluation of integrated interventions layered on mass drug administration for urogenital schistosomiasis elimination: a cluster-randomised trial. *The Lancet Global Health* 2019 Aug; 7:e1118–e1129. doi: 10.1016/s2214-109x(19)30189-5
27. World Health Organization. Field use of molluscicides in schistosomiasis control programmes: an operational manual for programme managers. Geneva: World Health Organization, 2017
28. Jordan P, Woodstock L, Unrau GO, and Cook JA. Control of *Schistosoma mansoni* transmission by provision of domestic water supplies: A preliminary report of a study in St Lucia. *Bulletin of the World Health Organization* 1975; 52:9–20
29. Jordan P. Some Quantitative Aspects of Bilharzia with particular reference to Suppressive Therapy and Mollusciciding in Control of *S. haematobium* in Sukumaland, Tanganyika. *East African Medical Journal* 1963 May; 40:250–60
30. Oettle RC. Human antibody-mediated immunity against infection and parasite fecundity in *Schistosoma mansoni* and *Schistosoma haematobium* infection. PhD thesis. University of Cambridge, 2020 Jul
31. Anyolitho MK, Poels K, Huyse T, Tumusiime J, Mugabi F, Tolo CU, Masquillier C, and Nyakato VN. Knowledge, attitudes, and practices regarding schistosomiasis infection and prevention: A mixed-methods study among endemic communities of western Uganda. *PLoS Neglected Tropical Diseases* 2022 Feb; 16. Ed. by Dikomitil L:e0010190. doi: 10.1371/journal.pntd.0010190
32. Sangare M, Diabate AF, Coulibaly YI, Tanapo D, Thera SO, Dolo H, Dicko I, Coulibaly O, Sall B, Traore F, Doumbia S, Kulkarni MA, Nutman TB, and Krentel A. 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 Global Health* 2024 Oct; 9:e015671. doi: 10.1136/bmjgh-2024-015671
33. Souza DK de, Gass K, Otchere J, Htet YM, Asiedu O, Marfo B, Biritwum NK, Boakye DA, and Ahorlu CS. Review of MDA registers for Lymphatic Filariasis: Findings, and potential uses in addressing the endgame elimination challenges. *PLoS Neglected Tropical Diseases* 2020 May; 14:e0008306. doi: 10.1371/journal.pntd.0008306
34. Malizia V, Vlas SJ de, Roes KCB, and Giardina F. Revisiting the impact of *Schistosoma mansoni* regulating mechanisms on transmission dynamics using SchiSTOP, a novel modelling framework. *PLoS Neglected Tropical Diseases* 2024 Sep; 18:e0012464. doi: 10.1371/journal.pntd.0012464
35. Kloos H, Gazzinelli A, and Van Zuyle P. Microgeographical patterns of schistosomiasis and water contact behavior; examples from Africa and Brazil. *Memórias do Instituto Oswaldo Cruz* 1998; 93:37–50. doi: 10.1590/s0074-02761998000700006
36. Kloos H, Fulford A, Butterworth A, Sturrock R, Ouma J, Kariuki H, Thiongo F, Dalton P, and Klumpp R. Spatial patterns of human water contact and *Schistosoma mansoni* transmission and infection in four rural areas in Machakos District, Kenya. *Social Science & Medicine* 1997 Apr; 44:949–68. doi: 10.1016/S0277-9536(96)00218-3
37. Meurs L, Mbow M, Boon N, Broeck F van den, Vereecken K, Dièye TN, Abatih E, Huyse T, Mboup S, and Polman K. Micro-Geographical Heterogeneity in *Schistosoma mansoni* and *S. haematobium* Infection and Morbidity in a Co-Endemic Community in Northern Senegal. *PLoS Neglected Tropical Diseases* 2013 Dec; 7:e2608. doi: 10.1371/journal.pntd.0002608

