Integrating Socio-Economic and Environmental Interventions to Improve Well-Being in Vulnerable Communities[1]

Pre-Analysis Plan

March 2025 amendment (original: April 2024)

Amendment note: This amendment reflects adjustments made after baseline data collection, which was completed in May 2024 after filing the original PAP, and community and household assignment to treatment, which likewise occurred after the original PAP (in May-June 2024). Thanks to receiving additional funding soon after treatment, we added additional data collection to cover an additional primary outcome concerning children's education, which we describe below. We also describe midline adjustments to data collection, including panel household tracing and replacement protocols. All content that is new with this amendment is in red text; deletions from the original PAP are in ~~strikethrough red text~~.

Problem statement[2]

Poor rural communities often lack sufficient food and clean water to maintain human health and productivity, and face a high burden of infectious diseases, generating reinforcing feedback that causes poverty-disease traps. In these settings, periodic drug treatments routinely fail to eliminate infectious diseases if they do not also address the disease's environmental reservoir; one needs to directly address the structural environmental mechanisms, not just the infections that are the symptom of environmental exposure. For example, in northern Senegal, the setting for this study, the prevalence of schistosomiasis (also known as bilharzia) in children often rebounds to 70-90% within a year after deworming drug treatment.

Schistosomiasis is the second most socioeconomically-burdensome parasitic disease globally, after malaria, affecting roughly 250 million people worldwide, with >800 million at risk and ~20 million suffering severe consequences annually. Schistosomiasis is caused by snail-transmitted flatworms (of the *Schistosoma* genus) that penetrate human skin. Even when provided drugs to clear the infections, humans quickly get re-infected when they return to snail-infested water bodies. Such persistent infection damages children's health and education advancement, and reinforces poverty. The disease has defied control efforts in the study region and most of the low-income tropics, and is prevalent throughout

This project studies a recent innovation that directly targets an environmental reservoir for the disease. Specifically, aquatic vegetation removal around water access points was recently shown to significantly reduce the burden of schistosomiasis in researcher-managed, pre-registered field trials (Rohr et al. *Nature* 2023). In this study, we explore the effectiveness of

[1] This project is funded by NSF (award numbers BCS-2307944 and DEB-2109293). The project principal investigators are Chris Barrett (Cornell), Nicolas Jouanard (Station d'Innovation Aquacole, SIA), Samba Mbaye (Centre de Recherche Pour Le Développement Économique Et Social, CRDES), Gilles Riveau (Espoir Pour La Santé, EPLS), Jason Rohr (Notre Dame, ND), Doudou Sow (Université Cheikh Anta Diop, UCAD), and Faraz Usmani (Mathematica), with key personnel including Sidy Bakhoum (ND), Molly Doruska (Cornell), Souleymane Doucouré (UGB), Lakshmi Iyer (ND), Kira Lancker (University of Copenhagen), Tidjani Ly (EPLS), Abdoulaye Mbengue (EPLS), Momy Seck (SIA), Bruno Senghor (UGB), SImon Senghor (EPLS), and Amina Sylla (CRDES).

[2] This section draws heavily on Rohr et al. (2023).

alternative designs for an information campaign (i) to promote adoption of that innovation and (ii) to stimulate improvements in schistosomiasis infection rates and living standards with local population-managed implementation of the innovation.

In our study region, a large majority of host snails are captured on or near the freshwater plant *Ceratophyllum demersum* (hereafter, *Cerato*). This plant (*i*) has a mutualistic relationship with snails, (*ii*) is found throughout Africa, Southeast Asia, and Latin America in areas where schistosomiasis is endemic, and, along with other invasive aquatic plants, (*iii*) chokes out waterways, impeding access to open water needed for washing clothes, irrigation, and cooking. Growth of these plants is stimulated by run-off of fertilizer and livestock manure into watersheds. Thus, agricultural development may inadvertently fuel infectious disease and hamper water access. The innovation developed and evaluated by Rohr et al. involves regular removal of *Cerato* to eliminate snail habitat and thereby reduce human schistosomiasis exposure.

The randomized controlled trials (RCTs) reported in Rohr et al. (2023) established not only the efficacy of aquatic vegetation, especially *Cerato,* removal (CR) in reducing schistosomiasis prevalence, but also the profitability of using the harvested *Cerato* as feedstock for compost applied to onion and pepper plots, the cost-effectiveness of its use as livestock feed—when dried for an adequate period of time to kill prospective parasites and pathogens—as well as the absence of significant unintended impacts on human water use or aquatic ecology. However, those results come from researcher-managed trials and thus are neither scalable nor sustainable unless local communities undertake CR on their own. The central objective of this study is to test among two different methods of extending information to try to induce manual CR by rural village residents, to see whether either or both intervention — individually or in combination—effectively induces CR and suppresses snail populations and schistosomiasis infection, improving living standards through any of multiple pathways. We also try to identify the specific mechanisms that generate any observed impacts and the distribution of such impacts within the population.

It is important to note that the snails that vector schistosomiasis are also hosted by other aquatic vegetation species besides *cerato* and even by debris such as used clothes and discarded plastic or wood. So general aquatic vegetation removal (AVR) is desirable to help reduce the vector habitat and reduce schistosomiasis exposure. Other aquatic vegetation can also serve as useful feedstock for compost production. But the researcher-managed trials reported in Rohr et al. (2023) focused on cerato so we emphasize CR specifically, and AVR more generally in the treatments described below.

CR is not especially time-consuming, but it does require regular effort, which necessarily diverts time that could otherwise be used for income generation, domestic chores, social activities, or leisure, all of which have value in poor rural communities. CR also involves some risk of infection if one does not use personal protective equipment (PPE).[3] For this reason, people need a good reason to engage in this innovative behavior.

CR for infectious disease control is a public good. Local and national governments do not presently provide this service. Private individuals must therefore be motivated to provide labor towards the public good. If people are solely self-interested, however, economic theory predicts that relying on voluntary private donation of costly and risky labor effort will result

[3] As described below, information experiment treatment arm participants were provided with chest waders, shoulder-length gloves, and pitchforks, along with instruction in why and how to properly use that PPE.

in suboptimal provisioning of the pure public good (CR), and thus a higher prevalence of schistosomiasis than is socially desirable. At the same time, if villagers also value public goods (such as children's health) and people are sufficiently pro-social, public health messages may suffice to control snail populations and limit disease prevalence by inducing the voluntary private provision of pure public goods. It is thus ultimately an empirical question whether simply explaining the public health benefits of CR will suffice to induce that novel behavior. Or perhaps people need to see some added, privately appropriate benefit from CR, as might be gained from the use of harvested aquatic biomass for compost or livestock feed, turning CR into an impure public good.

We designed an RCT to test information campaigns of the sort a government or non-governmental organization (NGO) might launch to promote manual CR by rural community residents. Specifically, we test whether communicating (*i*) the expected private agricultural productivity benefits from composted Cerato, (*ii*) the expected public health benefits from CR, or (*iii*) both induces CR and the follow-on benefits that Rohr et al. (2023) found in researcher-managed CR. This pre-analysis plan (PAP) describes the research design, our research questions (including both primary and secondary outcomes), our data collection methods, and our empirical strategy for testing the hypotheses in our research questions.

We hypothesize that:

- Communicating the private and/or public benefits of CR via an information campaign generates measurable CR, snail population reduction, and public health co-benefits that manifest in lower prevalence and severity of schistosomiasis infection;
- Educating farmers on the private benefits of CR—that is, an impure public good—induces increased labor effort in CR, relative to both a pure control group (that receives no information about CR) and an alternative information treatment arm that is only educated on the public health benefits of CR—that is, a pure public good;
- The private benefits treatment induces higher rates of compost use, leading to higher private agricultural productivity and incomes; and
- These benefits accrue disproportionately to poorer households, who are less likely to purchase fertilizer, have access to piped water (so as to otherwise minimize risks of infection through water contact), and who tend to have a lower opportunity cost of labor.
- The treatments affect children's school participation.

We also test whether encouraging CR for personal gain inadvertently reduces within-community cooperation or promotes individualistic behaviors over communitarian ones, generally and in the management of common pool resources (CPRs), such as the water sources and aquatic vegetation therein. For example, promoting individual seizure of CPRs may promote a more individualistic, Lockean perspective on resource tenure, reducing support for more communal, cooperative tenurial systems.

Finally, we monitor and test whether CR inadvertently disrupts aquatic ecology or water quality - relative to upstream and downstream control sites - and whether it induces increased human use of more accessible water; Rohr et al. (2023) found no such effects in the researcher-managed CR RCTs.

Background on the Senegal River Valley Region

This study takes place in Saint Louis and Louga regions of northern Senegal. The study communities are located in the Senegal River valley, adjacent to the Senegal River, Lac de Guiers or connected to irrigation canals that can host aquatic snails. Schistosomiasis has long been a major public health problem in this area, aggravated by aquatic ecology changes following the 1988 construction of the Diama Dam near Richard Toll (Southgate 1997, Diop et al. 2023). Two forms of schistosomiasis exist in this region: (i) *S. mansoni*, which infects the gastro-intestinal tract, and (ii) *S. haematobium*, which infects the urinary tract.[4] The statistically significant impacts identified by Rohr et al. (2023) were with respect to *S. mansoni* in particular.

Communities in this area are poor. Beyond the coastal city of Saint-Louis, few non-agricultural livelihood options exist, and most households depend heavily upon crop cultivation (mainly during the July–October rainy season) and livestock husbandry. Agricultural technologies in use are relatively rudimentary, with little mechanization. Crop yields and livestock lactation rates are very low by global standards.

Residents frequently rely on surface water to wash clothes, bathe, and collect water for cooking and drinking. Schistosomiasis prevalence in this area is therefore the highest of any region of Senegal (Diop et al. 2023). Since 2010, the national government has been running a schistosomiasis control program that includes regular deworming campaigns through schools in the region as well as preventative administration of deworming medication (typically praziquantel) among adults. However, the disease still constitutes a major health concern in this area, with prevalence rates among school children exceeding 87% (Léger et al., 2020; Senghor et al., 2022).

Research design

Overview

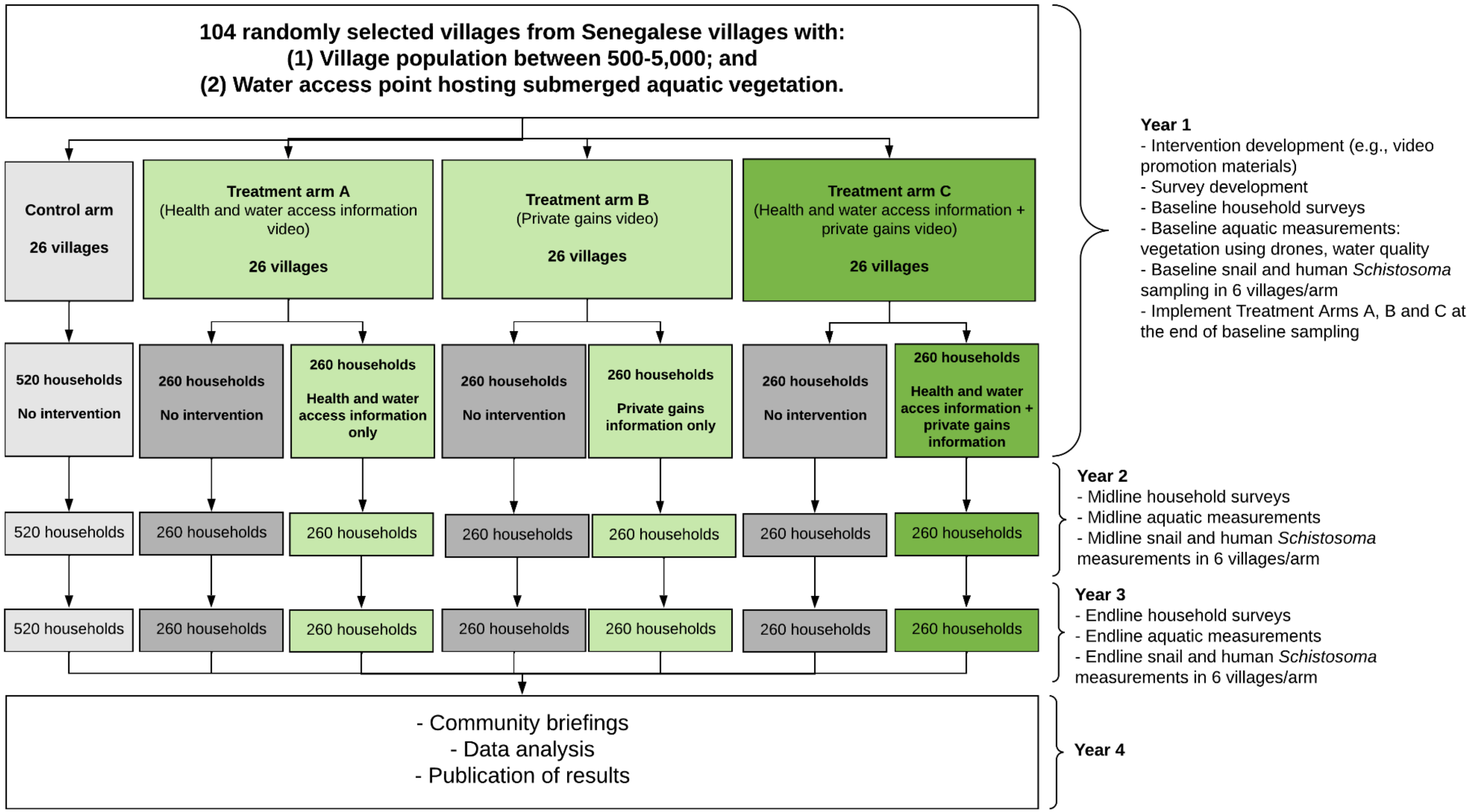
Our design consists of a cluster randomized 2×2 before-after control-intervention (BACI) trial (Figure 1). Specifically, we randomly divided 104 villages (originally, 88 villages, but we added 16 more, as explained below) into four arms of 26 villages each, including a control arm, and three treatment arms (arms A, B and C). Within each village, we randomly select and recruit 20 households for participation in the study, resulting in a total of 520 households in each of the study arms, for a total of 2,080 survey households. Within each treatment village, we will split selected households into 10 households who will not be directly exposed to the intervention and 10 households who will be invited to participate in the intervention. We refer to households in control arm of the study—that is, the 26 villages in the control arm that do not receive any intervention whatsoever, in line with the status quo scenario—as the “pure controls,” and to the 10 households per treatment village who are not be directly exposed to the intervention within treatment arms A–C as “local controls.”

Description of the intervention

Our intervention entails a roughly two-hour information session delivered to 10 randomly selected households in each village in the three information treatment arms (arms A, B and C). The information session consists of a standardized educational video - produced and delivered in the local languages, Wolof and Pulaar – that describes the water-access and schistosomiasis-reduction benefits of vegetation removal (“public health benefits”) or the

[4] *Schistosoma bovis* also infects ruminant livestock in the area and has been hybridizing with *S. mansoni* and *S. haematobium*, but remains unconfirmed in humans.

Figure 1: Intervention design



crop productivity and profit benefits of vegetation removal ("private benefits"), respectively, in treatment arms A and B. Both educational videos are shown to participants in the third treatment (arm C), thereby combining the public health and private benefits information treatments to create a full 2x2 BACI design. Each training video also includes instruction about appropriate precautions to take to protect oneself from infection when clearing vegetation by wearing personal protection equipment (PPE). Participants are given an opportunity and trained in how to properly don the PPE during the session. In addition, those receiving the private benefits information session are also trained on how to effectively convert the vegetation to compost and use the compost for crop production.

In addition to the educational video, experts will be present to answer questions and foster discussion among attendees and a local farmer with experience using compost created from CR will be present to attest to the benefits in the private benefits arm, and a public health expert will attend the public benefits arm to answer questions and foster discussion among attendees. We will also provide two sets of personal protective equipment (namely, a pitchfork, chest waders with boots, and full-length gloves) to be shared among each group of 10 attendees in each information session. Lastly, we will give each information treatment participant a short questionnaire to assess understanding of the benefits, risks and methods of harvesting aquatic vegetation, use for compost (if applicable), and personal protection. Before they depart the training session, each participant is provided with a laminated handout to be taken home to remind them of the value of aquatic vegetation removal. We also follow up with monthly reminders via mobile phone messages for one year after the treatment, conveyed through the village *relais communautaires* (relays) - community contacts established for a range of purposes for communicating with government and outside nongovernmental agencies – or another individual designated by the group of 10 participants at the time of training. Each of the relays is given air time credit of FCFA5,000 (just over US$8) each month to cover their messaging costs. At endline, we will share information on both the private and public benefits with all sample households.

We collect several different types of data: household surveys, community surveys and focus group discussions, drone imaging to measure the extent of vegetation removal, water sampling to assess the presence of snails, and urine and stool samples to assess schistosomiasis prevalence among school children. Beginning with the midline data collection, we also conduct school-based data collection in each survey village's main primary school. This involves a survey of each school's director/principal and direct in-classroom verification of attendance status of all children enrolled in the study. The data collection details are described below.

Village selection took place in November-December 2023. Household selection and baseline surveys began in January 2024, and concluded in ~~February~~ April 2024, with an interruption due to Ramadan. Ecological data collection and schoolchildren stool and urine collection and testing began in December 2023 and concluded in ~~early March~~ April 2024. At baseline, each household also participated in a pair of donation games. In addition, focus group discussions were held in each village at baseline with 6-10 participants not included in the baseline survey sample.

Delivery of the intervention is expected to start in mid-April 2024 and take 2-3 weeks to complete as shown in Figure 1. (Note: it began in late April and run until early June 2024.) We plan to follow sample households for two additional years with midline and endline surveys (in years two and three, respectively), supplemented with semi-annual drone imagery and net sweeps to quantify open water, snail populations, human water contact patterns and submerged vegetation in each water access point, and semi-annual school children attendance checks starting in 2025. We will repeat the donation games and focus group discussions at endline.

Research questions

In this section, we describe our main research questions, associated outcomes and, where relevant, key hypotheses. We group closely related research questions by the level at which associated outcomes will be measured and thematic focus.

1. **Primary outcomes**

   1.1. **Household- or individual-level:**

   Diffusion of CR practices: Cerato removal is the hypothesized mechanism through which beneficial results arise from the experiment. Accordingly, a primary outcome of interest—logically precedent to the others—is whether the information treatments indeed induce CR—or aquatic vegetation removal (AVR), more broadly since people may have difficulty identifying *cerato* reliably apart from other aquatic vegetation species and other aquatic vegetation can and does host the snails that vector *schistosoma*.

   1.1.1. **Does training induce AVR (measured by self-reports)? Does the AVR response to private benefits information differ from that to public health benefits information, versus information on both types of benefits together, all as compared to pure controls that receive no information?** Such responses are the initial mechanism we hypothesize leads to improved health and living standards.

   1.1.2. **Does training spill over to non-treated villagers (local controls) to induce them to engage in AVR? Does local spillover AVR response to information about private agricultural benefits differ in its adoption spillovers, versus information about public health benefits, versus information on both types of benefits together, all compared to pure control villages?** The policy-relevant aspiration is that training a subset of villagers suffices to spread the word and engage others in AVR.

   1.1.3. **Do we observe no uptake of AVR in pure control villages from baseline to endline?** One threat to identification of a causal effect of the information treatments (in 1.1.1 and 1.1.2) is the possibility that AVR spreads to pure control villages as well. As widespread diffusion of AVR can be considered a desirable outcome from a policy perspective—even if it might confound causal identification under our research design—we include this hypothesis. At the same time, engaging in AVR without appropriate protective equipment can increase risk of infection. We therefore aim to minimize spillovers (for instance, by ensuring that sample communities are not located too near to each other).

   1.1.4. **Conditional on finding AVR, does uptake increase between midline and endline, i.e., does the diffusion of AVR accelerate?** Diffusion of innovations typically follows an S-shaped curve in time, accelerating in early years before tapering towards steady state uptake levels. Does this intervention induce the apparent start of such a pattern?

Increased compost use, improved agricultural productivity and food security: The private benefits treatment arm provides simple, video-based training on how to make and apply compost created from harvested *cerato* and explains the evidence on the profitability of this practice. We seek to establish whether the training worked to induce uptake of compost production or use by trainees, as well as spillover to non-trainees. Trainees might be induced to directly produce compost. Or they might be induced to buy compost from those induced to produce it. Compost production and use could directly generate agricultural productivity gains. Note, however, that the treatment could also indirectly generate agricultural productivity gains through other channels, such as other uses of harvested *cerato* (e.g., as animal feed) or by improving the health of family members, thereby boosting labor supply and productivity. We cannot fully disentangle the direct and indirect pathways through which induced AVR increases compost use and agricultural productivity.

1.1.5. **Does training on the private benefits of CR induce compost production by treated households? Does training on the private benefits of CR induce compost use by treated households, whether through own production or purchase? Does the content of the training matter, or might inducing CR prompt composting even without compost-related messaging (i.e., for households with only the public health information treatment)?** We will compare against local controls and against pure controls to establish whether there is an effect of training regardless of its specific content.

1.1.6. **Does training on the private benefits of compost from CR spill over to non-treated neighbors (i.e., local controls) to induce them to engage in CR and compost production? Does that effect emerge as well in villages with public health benefits information treatments?** We hypothesize that information spillover is less when the messaging emphasizes private benefits, as trainees will be less likely to promote CR among neighbors with whom they might then compete for compost. This spillover mechanism balances out the incentive advantages of the private benefits information treatment over the public health information treatment since the latter is vulnerable to free riding problems.

1.1.7. Does training on the private benefits of compost from CR cause increased agricultural total factor productivity (value of total output divided by value of all inputs) and profitability? Does that effect also emerge in villages with only public health benefits information treatments? Are those effects greatest for poorer households, who are ex ante less likely to invest in chemical fertilizers and other improved inputs?

1.1.8. **Does training on the private benefits of CR and its use in compost production boost food security (as reflected in reduced self-reported months of food insecurity – known locally as *soudure* – and reduced coping strategies)?** We hypothesize that the gains will be greatest among poorer households because they are less likely to purchase chemical fertilizers and more vulnerable to schistosomiasis infections as they often lack access to piped water at baseline.

Reduced schistosomiasis: One of the target outcomes of the intervention—mediated through AVR (specifically, CR)—is reduced schistosomiasis prevalence and intensity (i.e., egg counts in stool or in urine).

1.1.9. **Does training in a village reduce the prevalence of schistosomiasis infection (from self-reported condition and symptoms, as well as from urine and stool sample testing among school children) as we compare treatment village households with pure control households? Does being in a village trained on the private benefits of CR yield greater reduction in schistosomiasis than being trained on the public health benefits only, presumably because of reduced free riding? Does being trained oneself reduce the prevalence of schistosomiasis, as we compare treatment participants versus local controls (can only test in self-reported data)?** Conditional on finding that training induced AVR, we hypothesize that differences with pure controls will be significant, differences between treated and local controls insignificant due to public health spillover benefits, and differences between private benefits and public health information treatments will be insignificant because the greater incentive effect of the private benefits information gets offset by how it attenuates treated individuals' propensity to share information on the benefits of CR. Note that both the private benefits and public health benefits training emphasize the disease risk of schistosomiasis exposure through unprotected human water contact and promote the use of PPE (which we provided). So it seems unlikely that any such differences would emerge because one treatment arm is differentially discouraged from entering the water unprotected.

1.1.10. **Does training in a village reduce the severity of schistosomiasis infection conditional on infection (from urine and stool sample testing among school children) as we compare treatment village households with pure control households? Does being in a village trained on the private benefits of CR yield greater reduction in schistosomiasis egg loads than being trained on the public health benefits only, presumably because of reduced free riding?** Conditional on finding that training induced AVR, we hypothesize that differences with pure controls will be significant, differences between treated and local controls insignificant due to public health spillover benefits, and differences between private benefits and public health information treatments will be insignificant because the greater incentive effect of the private benefits information gets offset by how it attenuates treated individuals' propensity to share information on the benefits of CR.

Pro-social behavior and property rights: The private benefits treatment encourages individuals to take individual possession of vegetation that is, in its natural state, a common pool resource (CPR). One might be concerned that this will encourage more individualistic behavior, manifest in greater support for Lockean conceptions of natural resource tenure (i.e., mixing one's labor with what was common property makes that resource one's own) and reduced willingness to contribute to the public good (as reflected in the donation games).

1.1.11. **Does the pre-intervention level of prosociality predict an individual's contribution to AVR? Do the information interventions affect contributions in the donation game? Do such effects spill over from treated households to local controls? How does an individual's propensity to donate relate to the individual's and the community's observable characteristics?** We hypothesize that individuals who contribute more in the donation game, and who are more prosocial as measured by Lockean beliefs in the household survey, are also more likely to contribute to AVR under treatments with public health benefit information (arms A and C), and that treatments that provide information on private benefits will decrease pro-sociality, as measured by donation game contributions. Further, we hypothesize that village level contributions are lower in villages with strongly perceived within-village inequality and individualistic beliefs, as obtained qualitatively from the focus group discussions.

1.1.12. **Does promoting the private benefits of a common pool resource (aquatic vegetation) induce a change in beliefs about property rights?** We hypothesize that the private benefits treatment will induce stronger beliefs in private property rights at the endline as measured by the beliefs module of the household survey, and as compared to pure controls and households in the public-only treatment arm.

Children's education: By affecting children's health status and potentially affecting household incomes we anticipate impacts on children's school attendance and performance in school conditional on attendance.

1.1.13 **Does training in a village change children's school participation (as observed at the school for primary-school-aged children present in study households at baseline)?** Competing mechanisms lead to an ambiguous prediction on potential impacts. On the one hand, improved health due to a reduction in schistosomiasis infections may improve school participation and hence educational attainment. On the other hand, the intervention also increases the opportunity cost of schooling, directly with CR as a new source of labor demand and indirectly as improved health also increases returns for other types of child labor, both of which may decrease school participation and hence educational attainment. Therefore, we do not have an explicit hypothesized impact of the intervention on child educational outcomes.

1.2. Water access point-level:

<u>**Reduced aquatic vegetation and snails in water access points:**</u> The purpose of the information treatments is to induce AVR. Self-reports of AVR help us understand if sample individuals (trainees or controls) engage in AVR directly. But the possibility of independent behavior by other, non-sample villagers could introduce a divergence between individual behavior and the state of the water access point. For example, trained individuals could encourage other, non-sample neighbors to clear aquatic vegetation, yielding the same village-level public health benefit as if the trainee cleared the vegetation themself.

1.2.1. **Does promoting the benefits of AVR reduce aquatic vegetation?** Using both drone imagery and manual net sweeps, we can observe whether greater AVR occurs in villages receiving both public and private benefits education relative to either one alone. We expect to see greater AVR in villages receiving education on the public or the private benefits education than in villages receiving no education at all. We will test this hypothesis by using two different measures. One is water access point level based on manual dip net sweeps at each access point before the treatment arms are implemented, and semi-annually thereafter once the treatment arms have been implemented, through endline.The other measure is for all the village water access points, and out to 100 meters from those points, based on submerged cerato presence extracted through an algorithm from drone imagery.

1.2.2. **Does promoting the benefits of AVR reduce aquatic snail populations, in particular of snails infected with schistosomiasis?** We hypothesize that we will observe significant drops in snail densities in villages receiving both public and private benefits education relative to either one alone. We also expect to see greater drops in snails densities in villages receiving education on the public or the private benefits education than villages receiving no education at all. We will test this hypothesis by using standardized dipnet sampling of snails at each water access point at villages before the treatment arms are implemented, and semi-annually at midline and endline after the treatment arms have been implemented. We test for schistosomiasis infection in snails by having the snails shed in controlled laboratory conditions the same day after dipnet capture.

2. Secondary outcomes

2.1. Household-level:

2.1.1. **Does training in a village reduce individuals' number of days of work or school lost due to ill health (from self-reported conditions and symptoms)?** This would draw together multiple mechanisms, through direct reduction in schistosomiasis exposure due to CR, indirect advances due to increased household incomes from reduced time lost to illness and improved agricultural productivity. But it can be confounded by a variety of external changes that could spuriously correlate with treatment. In addition, self-reported health measures are noisy. For this reason, we treat this as a secondary outcome. As with primary outcome 1.1.8, we will also test whether being trained oneself (i.e., trainees only, as compared to local controls) reduces the prevalence of self-reported illness, particularly in terms of days of school or work lost to the household. Conditional on finding that training induced AVR, we hypothesize that differences with pure controls will be significant, but differences between treated and local control households will be insignificant due to public health spillover benefits.

2.1.2. **Does training in a village change children's school participation and educational attainment (from self-reported measures on school-aged individuals)?** Competing mechanisms lead to an ambiguous prediction on potential impacts. On the one hand, improved health due to a reduction in schistosomiasis infections may improve school participation and hence educational attainment. On the other hand, the intervention also increases the opportunity cost of schooling, directly with CR as a new source of labor demand and indirectly as improved health also increases returns for other types of child labor, both of which may decrease school participation and hence educational attainment. Therefore, we do not have an explicit hypothesized impact of the intervention on child educational outcomes.

2.1.3. **Do individuals change their contributions when a pure public good is turned into an impure public good?** The addition of private gains when contributing to a public good (turning it into an impure public good) may reduce public contributions due to crowding out (Engelmann et al. 2017, Munro & Valente 2016, Guo et al. 2021) or anchoring, which is of interest for the effective design of information policies. Alternatively, the private benefit framing may change how the community benefits are viewed and may induce increased donations if it results in respondents feeling like they have more "skin in the game." Respondents who contribute less than CFA 200 (very few in our pilots) would likely increase their contributions. Our RCT would enable testing of such mechanisms only via cross-village comparisons; embedding both types of donation games within the survey allows us to test this using a within-individual design.

2.2. **Water access point-level:**

2.2.1. **Does training on the benefits from AVR induce change in human water use patterns?** We expect that sites with less vegetation obstructing water access might be more inviting for swimming and thus there might be an increase in water contact. However, we did not detect this in Rohr et al. (2023). Additionally, encouraging people to remove the vegetation might increase their water contact rates, despite providing personal protective equipment (PPE) if many villagers choose not to wear the PPE. We will test this hypothesis separately for pre-school age children, school-age children, and adults, using the counts of people in water from each semi-annual water access point data collection round.

2.2.2. **Does training on the benefits from AVR induce change in snail populations and aquatic vegetation (especially cerato) density?** We expect that our information treatments will induce increased AVR, which will manifest in both lower volume of submerged vegetation that creates habitat for snails as well as in lower snail populations.

2.3. Community-scale:

2.3.1. **Do information treatments induce changes in natural resource tenure of aquatic vegetation and/ or other, unrelated common pool resources?** We hypothesize that we will observe differences between villages of different treatment arms regarding changes in natural common pool resource tenure and management at village level, as per qualitative insights from focus group discussions and quantitative indicators from the community level survey.

2.3.2. **Do information treatments affect the prevalence and/or severity of schistosomiasis infections among schoolchildren?** Using the fecal and urine samples collected from 24 of the sample villages, we will test for differences among villages with (i) private benefits treatments,

(ii) public health benefits treatments, and (iii) pure controls in the prevalence and average worm count (infection load) per child.

2.3.3. **Do information treatments cause unintended effects on water quality or aquatic biodiversity, using upstream and downstream monitoring sites as controls?** Although Rohr et al. (2023) did not find significant effects of the CR on water quality or non-target organisms, increasing the scale of this intervention could result in unintended consequences not found in the initial trials. We will measure water quality and aquatic biodiversity at villages both upstream and downstream of villages enrolled in treatment arms to identify ecosystem-level effects of CR. We expect that up and downstream sites will not significantly differ in these variables if there are no substantial unintended consequences of CR on the ecosystem.

Power calculations

We present illustrative power calculations for different types of outcome variables and analyses in Table 1. Note that these power calculations do not account for corrections related to multiple outcome and multiple hypothesis testing that we will conduct, as described further below

Illustrative outcome variable	**Minimum detectable effect (units of outcome)**		
	Treatment vs. control arms Cluster-level randomization N = 2,080 households across 78 treatment villages and 26 control villages	**Across any two treatment arms** Cluster-level randomization N = 1,040 households across 26 treatment arm 1 villages and 26 treatment arm 2 villages	**Treated households vs. local controls within all treatment arms** Individual-level randomization N = 1,560 households of which 780 are treated and 780 are local controls
<u>Binary variable:</u> “Self-reported aquatic vegetation removal”	0.019 Assumed control group mean: 0.01	0.102 Assumed treatment arm 2 mean: 0.25	0.045 Assumed local control mean: 0.13
<u>Continuous variable:</u> “Number of months of *soudure* in past 12 months”	0.095 Assumed control group mean (SD): 3 (0.5)	0.070 Assumed treatment arm 2 mean (SD): 2.5 (0.3)	0.054 Assumed local control mean (SD): 2.75 (0.4)

Table 1: Illustrative power calculations

Notes: All power calculations assume a two-sided test, 0.05 significance level, and 80 percent power. Cluster-level randomization power calculations assume an intraclass correlation coefficient of 0.05, and the proportion of the within-cluster as well as cluster-level variance of the outcome explained by covariates equal to 0.10. Individual-level randomization power calculations assume that the proportion of the individual-level variance of the outcome explained by covariates is equal to 0.10.

Village selection

We initially randomly drew 88 villages that contain or are adjacent to a body of freshwater that could host submerged vegetation, such as *C. demersum*, and thereby serve as a reservoir for *Schistosoma*. We drew on village locations from the 2013 national census and existing GIS data from Google Earth Engine on surface water throughout Senegal to identify villages that met our criteria. We stratified villages based on the baseline agricultural intensity of the lands surrounding the village—as manifest in NDVI—as that influences nutrient runoff and thus *C. demersum* growth and baseline exposure to the disease. We then randomly sampled villages within the two strata to obtain our final sample of villages. We added 16 more villages to baseline at the last minute, as explained below, yielding a total of 104 villages, following exactly the same inclusion criteria and stratification and buffering procedures.

More precisely, to create the randomized listing of villages, we first limited the set of villages considered for an initial site visit using 2013 census-based listing previously constructed by SIA. If a village was listed jointly with another village, both villages were included separately, since the field team had to verify if these are in fact two different villages. Villages in which the field team had previously conducted intervention research that directly or indirectly communicated any findings from Rohr et al. (2023) or Doruska et al. (2024) were initially disqualified from inclusion in the sample due to pre-baseline contamination.

We stratified villages into those with above median NDVI readings and below median NDVI readings since Rohr et al. (2023) found that snail and schistosomiasis prevalence is positively associated with agricultural development. This stratification ensures adequate distribution of villages among those with a higher likelihood of heavy versus lighter pre-treatment exposure to the disease. We randomized villages into the various treatment and control arms within each stratum.

Nine villages already monitored by EPLS in a parallel study (Cartobil, in collaboration with researchers at Stanford University) were pre-selected for inclusion as they were known to satisfy all inclusion criteria and not to have been contaminated through any sort of intervention; we first randomized these villages into the four different experimental arms. Based on the allocation of these 9 villages, we then reduce the set of villages eligible for the various arms of the experiment based on their proximity to the already selected and randomized villages.

We imposed a 5-kilometer buffer among sample villages. For any village assigned to the control arm, any other village within 5 km of the village must also be in the control arm and cannot be in any treatment arm. For villages in the Private Benefits arm, any other village within 5 km of the village must be in either the Private Benefits arm or the Private and Public Benefits arm and cannot be in the control arm or the Public Benefits arm. For villages in the Public Benefits arm, any other village within 5 km must be in the Public Benefits arm or the Private and Public Benefits arm and cannot be in the control arm or the Private Benefits arm. For villages in the Private and Public Benefits arm, any village within cannot be in the control arm. Thus, the randomization of the 9 pre-selected Stanford/Cartobil villages imposed some restrictions on the rest of the village randomization process.

After eliminating villages not eligible for certain treatment arms due to proximity to already-assigned villages, we randomized - using a computer random number generator - villages one by one across the different treatment arms within each NDVI-based stratum. After selecting a village, we referenced the list of villages within its 5 km buffers and updated which experimental control arms these nearby villages were eligible to join. We followed this process until we had a listing of 104 randomly selected villages across the four experimental arms, with two strata within each arm.

A field team comprised of representatives from the CRDES, ND and SIA teams visited each of the 104 villages to ensure they satisfied the inclusion criteria, in particular, the village size and likely presence of *C. demersum* or schistosomiasis, and to secure the village chief's consent to include the village in the survey. The field team eliminated multiple villages as they did not satisfy one or more of the sample inclusion criteria. No chief of an otherwise eligible village refused to have that village participate. The team also elicited from each chief the preferred use of funds generated through the donation game.

After confirming a village's inclusion in the final sample, the geocoordinates and name and telephone number of the village chief were recorded in a confidential file to facilitate follow-up contact and data collection visits.

During baseline ecological data collection, the ND team doing the dipnet sweep sampling of snails and aquatic vegetation noticed that quite a few sites lacked *C. demersum*, snails, or both. That unexpected absence threatened the research design, because if no *C. demersum* is present, then treatments designed to induce CR will necessarily have no effect on *C. demersum* and are much less likely to have any impact on snail populations, which would seem to have a non-cerato host.

We therefore quickly summarized the ecological data to be more precise about the prospective problem. We found that 32 sample villages had no *C. demersum*, no snails, or neither *C. demersum* nor snails. Furthermore, those absences were not balanced across the four arms of the experiment. There is some reasonable chance that some of these sites experience purely seasonal *C. demersum* or snail absences such that once the rainy season begins (typically in July), *C. demersum* and snails will return. It is also possible - but less likely – that because the team only sampled one water access point per village, *C. demersum* and/or snails may have been present at one or more other (less-used) water access points used by that village, such that the null results reflect not seasonality but sampling error. In the case of either seasonality or sampling error, these sites remain valid and the experiment and hypothesized mechanisms remain relevant.

It seemed unlikely, however, that all 32 sites' snail or *C. demersum* absences were attributable to just sampling error or seasonality. More likely, schistosomiasis is present in those villages through some other transmission mechanism not targeted by our intervention. (Our team was collectively unaware of any village in the study region that had been screened for schistosomiasis and found to have zero prevalence in the last decade or more.) Most likely, some of these villages - our estimate was perhaps one-third – were erroneously included in the original sample. Their inclusion risks (i) significant attenuation bias in our estimates, and (ii) downward bias in the estimated (positive) impacts of the information interventions, especially with respect to the public health benefits information treatments (arms 1 and 3) in which we found the highest prevalence of zero-valued baseline observations for *C. demersum* or snails.

We therefore agreed to several corrective measures pre-intervention. First, starting with the July-August 2024 ecological sampling, we will cover up to two water access points per village - the two points most used by village residents, prioritizing those with *C. demersum* present - in the dipnet sweeps. The drone imagery will cover all water access points used by the villagers. Second, we re-randomized the 32 villages found to have no *C. demersum* or no snails so as to balance them across experimental arms. That requires reallocating 3 from treatment arm 1 to control, and 1 each from treatment arms 1 and 3 to treatment arm 2. Third, we added 16 villages to the sample, unequally across experimental arms so as to restore equal sample sizes across each arm after the re-randomization. Of these, eight villages had been originally excluded because they were controls in the Rohr et al. (2023) study and included in the Doruska et al. (2024) auctions. (As indicated below, we include an indicator variable for those villages in regressions.) Those 16 additional baseline surveys and ecological data collection were all completed in March-April 2024 prior to the information treatments. EPLS collected baseline stool and urine sample data from (27-30) school children in five of those villages, which augments that sub-sample, yielding a total of 29 villages from whose school children we collected stool and urine samples annually, starting with baseline.

The final village listing for the 104 villages, along with 12 upstream and downstream water quality monitoring sites, is shown in Appendix A.

Data collection

This section provides an overview of each of the data collection efforts conducted as part of this study.

Household- and community-level data collection

Household- and community-level data collection activities are being led by a team from the Centre de Recherche pour le Développement Économique et Social (CRDES). Prior to launching data collection activities, we trained and organized four survey teams, each consisting of one supervisor and four other enumerators. Training occurred from January 4–9, 2024 at Gaston Berger University, and included a one-day field pilot in the village of Ndiawdoune.

Data collection within sample villages started in January 2024, and concluded in mid-April 2024, just prior to the information treatments. Upon arriving in each village, survey teams first sought permission from the village chief to initiate data collection activities. After receiving permission, teams worked with the village chief to develop a roster of all households within the village along with the village chief's assessment of the household's relative wealth standing ("high" or "low") within the community, following which the village chief—or another community leader—completed a detailed community questionnaire to collect information on community-level characteristics (such as infrastructure availability, agricultural practices, and local prices).

A total of 20 households were then randomly selected from the village roster, stratified on relative wealth levels, for a total sample of 1,760 households. Randomly selected households were invited to complete a household questionnaire, which included modules to collect information on household composition and time use, health status (including knowledge about and incidence of schistosomiasis), income and living standards, agricultural practices, and beliefs and perceptions relating to individual and communal property rights.

Finally, households were invited to participate in two separate donation games. Specifically, households completed the following games, with the order in which the games were presented to the respondent randomized at the individual level:

- *Standard donations game*: Before the game starts, each participant receives an envelope with CFA 1,200 (one CFA 500 note and seven CFA 100 coins).[5] The enumerator reads the script to the participant (see Appendix C for all survey materials). The script states that respondents should divide up their CFA 1,200 in one part to keep for their own use (private) and a second part to donate for the community gift (public contribution) to the village-serving organization previously chosen by the village chief (either the local mosque, health facility, or school). Individuals' public contributions are noted down by the game coordinator. The game coordinator stresses that aggregate public contributions, after the household surveys are finalized in the village, will be increased by 50 percent by the survey team and donated to the pre-designated community gift in a public ceremony at the end of the research team's visit to the village. The enumerator gives the participant the time and place of that gathering, helping instill trust in participants that their contribution to the community gift will actually reach its destination safely.
- *"Impure" donations game*: This variant of the game changes the incentives for the donation contribution relative to the standard donation game. First, the initial endowment is CFA 1,000 (one CFA 500 note and five CFA 100 coins). For the first CFA 200 contributed to the public good ("threshold"), the respondents unconditionally obtain an individual benefit of CFA 200, that is, if they donate at least CFA 200, they will be given an additional CFA 200 on top of the initial CFA 1,000 endowment. All other aspects of the game and how it is administered are unchanged. This means that respondents who would contribute CFA 200 or more in the standard donation game will have no monetary incentive to change their contributions. Comparing the contributions between these two variants of the game will enable estimation of any behavioral mechanisms induced by the presence of private benefits.

Starting with the midline data collection that began in January 2025, the household survey team implemented the following household tracking and replacement protocol to ensure maximal retention of baseline survey households and representativity of the survey villages from the baseline period.

Enumerators would revisit all households surveyed at baseline. The enumerator would verify with the respondent that they had indeed been surveyed the prior year. Baseline data would then get imported on the SurveyCTO CAPI and the enumerator would confirm household roster members. For any household that the enumerator could not initially reach, first the enumerator would attempt to make contact via telephone to determine the household's whereabouts. If that failed, the enumerator would attempt to identify the household's location via the village chief. If the household was still in the village and the original respondent was unavailable, another adult household member was recruited to respond on behalf of the household. If the household was temporarily away from the study village but would return during the midline data collection period, they were to be revisited later. If the household would not return to the study village during the midline data collection period, they were to be replaced by the next replacement household from the same village from the replaced

[5] Due to a shortage of small denomination notes and coins, participants were paid via mobile money in a subset of surveyed villages.

household's baseline wealth stratum. A slightly modified survey instrument was used for replacement households so as to ensure capture of all baseline data not being collected again at midline from repeat respondents, in addition to the midline data.

School-based data collection:

We will implement school-based data collection at midline and endline at each community's main primary school. Most communities have one primary school located in that community. In cases where communities do not have a primary school (~6 communities), we will collect data at the primary school identified by the village chief as the "main" primary school for that community. In cases where communities have two primary schools (~6 communities), we will collect data at both primary schools, inquiring on all study children at one primary school and then inquiring at the second primary school on the subset of study children not enrolled in the first primary school.

At midline, enumerators will interview school directors/principals and collect data on school director/principal characteristics, school characteristics, and school enrollment by grade level for the current and previous school year. Then, enumerators will use the school roster book to verify the enrollment status of all primary-school-aged children present in study households at baseline and, if enrolled, their current grade level and classroom. Finally, enumerators will go to each classroom and ask the teacher to identify the attendance status of all enrolled study children. There is no direct interaction with the children.

Enumerators will also perform another attendance check in between midline and endline, which we call "post-midline". This will occur roughly five to eight months after midline data collection, after the rainy season is well established – a period when schistosomiasis infections typically increase – to roughly correspond with or follow soon after the study's ecological sampling in July-August 2025. First, enumerators will record any changes to school director/principal characteristics, school characteristics, and school enrollment by grade level (e.g., if a school director has changed). If the new school year has started, enumerators will use the school roster book to verify the enrollment status of all primary-school-aged children present in study households at baseline and, if enrolled, their current grade level and classroom. Then, enumerators will go to each classroom and ask the teacher to identify the attendance status of all study children who are known to be enrolled in that classroom at the time.

Finally, at endline, enumerators will interview school directors/principals in the next school year and update data on school director/principal characteristics, school characteristics, and school enrollment by grade level. Then, enumerators will verify the enrollment status and attendance status of all primary-school-aged children present in study households at baseline, following the same procedure as implemented at midline.

Focus group discussions

Baseline focus group discussions started in January 2024 in conjunction with the household surveys, and concluded in early April 2024. We conducted focus group discussions regarding tenurial control over resources, as well as well-being and health dynamics. In each village of all four treatment arms, an open discussion along a catalog of 17 open-ended questions was held with 6-10 adult, non-survey participants. Participants were selected according to the following criteria: all participants were selected from different families and had to be fluent in Wolof, over 18 years old, and in good health. To ensure diversity, we chose at least two men and two women, at least two participants younger than 40 and two older than 40, at least one participant from the lower and one from the higher end of the wealth distribution and ideally, participants from different parts of the village.

Ecological (sweeps, drone) sampling and measurement protocols

A team from ND and SIA began baseline data collection in December 2023 and concluded data collection in early April 2024. In each village, we sampled the water access point most used by village residents. The drone flights were done by SIA at the same water access points from which sweep samples were collected by a ND team.

The ND team that did the dipnet sweep sampling also gathered data on environmental factor predictors of snail abundance. At baseline, they selected one water access point per village. During the November-December previsit, we asked how many water points villagers used, and the team then went to manually inspect each of them. If there is more than one access point in the village (where access points are defined by emergent vegetation on either side), we asked first the biggest and most frequented access point, and if it had any cerato, we sampled that point. If the most used point did not have cerato, we sampled the most used point that did have cerato. If no cerato was present - which could be simply a seasonal phenomenon since we did baseline sampling well into the dry season – we sampled the most used water access point for water chemistry, vegetation, snails, and *Schistosoma* parasites in snails. Drone flights were conducted at every water access point at each village to estimate submerged vegetation at village scale. So we have two distinct measures of submerged vegetation presence: one at water access point level based on dipnet sweep samples, the other at village level based on machine learning-based estimates from drone imagery (for details, see Appendix ~~C~~B).

Starting with the midline sampling, we began sampling across the two main water access points used by each community, which were determined by direct communication with the village chief and verified by the sampling team. Many villages had only one main water access point, but 38 villages had 2 access points. Water access points were sampled regardless of cerato presence or hydrological feature (river, lake, or canal). Drone flights were conducted at every water access point at each village to estimate submerged vegetation at the village scale.

At each sampled water access point, the team recorded pH, water conductivity, water temperature, salinity and total dissolved solids (TDS) using a YSI Professional Plus handheld multiparameter meter. We collected a phytoplankton sample in undisturbed water by filling a 15-ml plastic sample tube. We cut across Typha or other emergent vegetation at the water surface with scissors, then inserted the top end into a 50-ml sample tube. We cut the bottom end clean at the tube opening. We kept periphyton and phytoplankton tubes in the dark for one hour before testing in the lab. In the lab, we filled the 50-ml sample tube containing Typha with 45 ml of water and removed all the periphyton with a toothbrush, rinsing the brush in the tube to remove followed by vigorous shaking. Then, we took an aliquot of periphyton using a pipette to half fill a fluorometer cuvette tube. We used the fluorometer to record Ft and QY values on the datasheet for periphyton and phytoplankton using the cuvettes. We rinsed cuvettes with water. We recorded the length and width/diameter of the clipping using a caliper in the datasheet.

During baseline sampling, at each access point, we performed 10 1-m dipnet sweeps within the boundaries of the water point: three open and seven submerged (on the *Cerato,* if present). Some villages, especially further east – in the Podor and Ndioum areas – lacked emergent vegetation delimiting access points; these were basically beaches along the river, so sweeps were just conducted along the shore at a common access point and separated by the same distance. In each sweep, we noted which microhabitat was swept in the datasheet. Captured plants were placed into a bucket with water, and shaken vigorously to remove snails and other animals before being examined for any remaining attached snails before being weighed using a digital hanging scale. If there was no *Cerato* in the sweep, other plants were weighed. We poured the water in the bucket through a strainer and collected snails into a pre-labeled sample container. We recorded the number of snails by genus and other animal groups per sweep in the datasheet, along with the sweep depth using a one-meter caliper as well as the GPS location of the sweep. We recorded the snail container number, phytoplankton and periphyton sample tube numbers on the datasheet for each access site and transport captured these back to the lab in a cooler until shed. At the few water access points where no vegetation was present, we performed sweeps on the debris found at the site (e.g., wood, used clothes, plastic, etc.) or on the open mud/sand.

During midline sampling at each access point, we performed three 1-m dipnet sweeps across three transects for a total of nine sweeps. The transects are used to standardize the data across water bodies of varying sizes, depths, and vegetation coverage. While transects may have to be adapted to a variety of different shapes and sizes of water access points, the goal is for there to always be a consistent distance between all sweeps and that sweeps span a reasonable extent of the access point. In general, each transect begins where the depth is about ankle height and a sweep is performed there. Then, the person sweeping moves directly perpendicularly out from shore to where the water is approximately knee-depth and performs a sweep there. Then, you move out perpendicularly again to where the water is approximately waist-depth and perform a sweep there. The next three sweeps across the transect should be parallel to the first but five meters in width from the first transect. If this width was adjusted based on the size of the water access point, this width was recorded. Three transects are performed, regardless of the presence of vegetation in any of the transect points, to standardize the sampling of each heterogeneous water body. For canals and water bodies of extensive length but minimal width, the transect can be adapted into nine equally spaced 1-m dipnet sweeps that are parallel rather than perpendicular to the shoreline. Vegetation should be recorded, *C. demersum* should be weighed, and any debris should be noted in each sweep.

All collected snails were brought to the laboratory the same day to determine if they were infected by *Schistosoma*. In the laboratory, individual snails were exposed under artificial light for one hour to promote schistosome cercarial shedding. Once cercariae were shed, Schistosomes were identified by their diagnostic forked tail and counted with the assistance of a dissecting microscope.

Infected snails are remotely sized by using Image J. They were placed first on a gird paper with known dimensions and photographed. Each start and stop time was noted in the datasheet. A count of all persons in contact with water (except people taking canoes to cross the river, and thus not making skin contact with water) was kept between the start and the end times of sampling. Starting with the first semi-annual follow-up round, we begin breaking down the human population in contact with water into (i) pre-school age children (apparently under or equal to five years old), (ii) school age children (roughly ~~5~~6-18 years old), and (iii) adults (seemingly over 18 years old).

The drone imagery data collection and analysis protocol can be found in Appendix B.

Parasitological sampling, testing and treatment

The EPLS team began baseline data collection in late November 2023 in 14 villages shared with another (Cartobil) project that is doing purely observational monitoring using the same sampling and testing protocol. That sampling concluded in February 2024. The UCAD/UGB team began baseline data collection in March 2024 in the other 15 villages in which stool and urine samples were collected from primary school children and tested. Their baseline was completed in April 2024, just prior to the information treatment interventions.

The sampling, testing and treatment protocols used were identical between EPLS and UCAD/UGB, using procedures developed already for an observational study (the Cartobil project) that EPLS was doing in collaboration with researchers from Stanford University. In each village, the research team received parental consent to sample (and treat, if their child was found infected) a target of 50 children enrolled in the local primary school. So as to maximize the likelihood of tracking of children over the three survey waves, and because schistosomiasis' effects are most acute among younger children, in every village the entire first year class was sampled. Conditional on parental consent, all children in the same classroom were sampled and treated, so as not to treat any child differently than their classmates. If there were not 50 students in the first year class, the team would also sample the second year class. If the first and second year classes together did not encompass 50 students, the team would sample the third year class, and so on until at least 50 primary school children were sampled or the full school child population of the village had been sampled, whichever came first. In many villages, the uniform treatment of students in a common classroom yielded more than 50 samples per school. In a few villages, the school has less than 50 children. So the per village samples are not uniformly 50 children.

A stool sample and a urine sample were collected from each child and analyzed in the laboratory on the same day to count Schistosoma sp eggs. The precise lab protocol for treating and analyzing samples and recording the results is standard, following Rohr et al. (2023). A second sample of both stool and urine were collected from each of the same children one week later. The second samples were analyzed only in the case of children whose first samples were negative (i.e., no Schistosoma sp eggs identified). The doubling sampling aims to minimize false negatives. In order to conserve scarce lab supplies, second samples were not analyzed in the case of children who tested positive in their first sample. The second sample was collected from those students anyway so as to maintain confidentiality of which children were found infected in the first sample. All sample children then received praziquantel to clear (and, for a period, prevent against) worm infections.

Each child's name, school year level, and parent name(s) were recorded. We use these to match children from the primary school sample with children in the household sample using a unique, child-specific identification code. That lets us link anonymized data sets.

The research teams coordinated in advance with the Ministry of Health to ensure that they did not include the survey schools in the annual (in principle) deworming campaign that typically begins in December. This was to ensure that children's infections were not cleared shortly before the research teams collected urine and stool samples for participating children. Specifically, we shared the study protocol with the coordinator of the national Neglected Tropical Diseases Control Program in Senegal to inform them about the study. We also engaged with the health district chief medical officer and then the list of the villages concerned was shared with the district and the directors. We asked them to not include these children in the mass drug administration efforts and committed to deworming the children after we completed our sampling that year. To ensure that children were not dewormed prior to sampling, the UCAD/UGB team participated in and helped supervise the Ministry's mass drug administration campaign in the field in this region.

After the two parasitological analyses spaced one week apart, all the children in the school were treated with praziquantel (deworming drug) a dose of 40 mg/kg and followed one year after treatment.

Empirical methods

Regression specifications

In this section, we present the regression specifications we will estimate to answer each research question (RQ) outlined in the Research Questions section above.

1. **Primary outcomes**

1.1. Household- or individual-level

Diffusion of CR practices

1.1.1. Does training induce AVR (measured by self-reports)?
Our analysis will focus primarily on intent-to-treat (ITT) effects of the intervention in villages in the treatment arms at midline and endline (examining each round separately). We will use analysis of covariance (ANCOVA) regression analysis to estimate impacts, conditioning on the baseline value of the relevant outcome variable to increase statistical power (McKenzie 2012). Because there may be spatial spillovers, we explicitly control for distance to the nearest village in a different treatment arm. Specifically, we will estimate the following general specification:

$$y_{iv} = \beta_0 + \beta_1 T_v + \beta_2{}' X_{iv} + \beta_3 y_{iv}^* + \gamma' D_v + \Theta' A_v + \epsilon_{iv} \qquad (1)$$

where y_{iv} is the outcome of interest for household i in village v at middle or endline; T is a binary variable that equals one if household i is located in a village randomly assigned to one of the three treatment arms, and zero otherwise; $X_{iv}^{'}$ includes controls for baseline village, household and/or individual characteristics, namely distance to nearest health clinic and number of water access points used by villagers (village-level variables), household size, access to piped water, and wealth as measured by a household asset index), and the household head's age, sex and literacy status (household-level variables); D_v is the four element vector of distance (in minutes walking to the nearest village in each of the four experimental arms, with a zero indicating the village is in that treatment arm); A is a dummy variable taking value one for villages that were in the Doruska et al. (2024) auctions experiment and zero otherwise, and y_{iv}^{*} is the baseline value of the outcome of interest. We will cluster standard errors at the village level in line with the village-level assignment of the treatment. If we find more than five percent of dependent variable observations are zero-valued, we will also estimate this (and other equations below) using a panel data censored dependent variable estimator (e.g., CLAD).

<u>Does the AVR response to private benefits information differ from that to public health benefits information, versus information on both types of benefits together, all as compared to pure controls that receive no information?</u>

We will estimate a modified version of the specification shown in equation (1), as follows:

$$y_{iv} = \beta_0 + \beta_1 T_A + \beta_2 T_B + \beta_3 T_C + X_{iv}^{'}\beta_4 + \beta_5 y_{iv}^{*} + \gamma' D_v + \theta' A_v + \epsilon_{iv} \qquad (2)$$

where T_A, T_B and T_C are binary variables that equal one if unit i is located in a village in treatment arms A, B or C, respectively, and zero otherwise.

1.1.2. <u>Does training spill over to non-treated villagers (local controls) to induce them to engage in AVR?</u>

We will measure within-village spillovers by disaggregating the different types of households and estimating the following modified version of equation (1):

$$y_{iv} = \beta_0 + \beta_1 T_i^L + \beta_2 T_i^T + X_{iv}^{'}\beta_3 + \beta_4 y_{iv}^{*} + \gamma' D_v + \theta' A_v + \epsilon_{iv} \qquad (3)$$

where T_i^L and T_i^T are binary variables that equal one if household i is a local control or treated household, respectively, in a village assigned to one of the three treatment arms.

<u>Does local spillover AVR response to information about private agricultural benefits differ in its adoption spillovers, versus information about public health benefits, versus information on both types of benefits together, all compared to pure control villages?</u>

We will disaggregate the different types of households and estimate the following modified version of specification shown in equation (2):

$$y_{iv} = \beta_0 + \beta_1 T_{iA}^L + \beta_2 T_{iA}^T + \beta_3 T_{iB}^L + \beta_4 T_{iB}^T + \beta_5 T_{iC}^L + \beta_6 T_{iC}^T + X_{iv}^{'}\beta_7 + \beta_8 y_{iv}^{*} + \gamma' D_v + \theta' A_v + \epsilon_{iv} \qquad (4)$$

where T_{iX}^L and T_{iX}^T are binary variables that equal one if household i is a local control or treated household, respectively, within a village in treatment arm $J \in \{A, B, C\}$.

1.1.3. <u>Do we observe no uptake of AVR in pure control villages from baseline to endline?</u>

We will conduct descriptive "before–after" analyses of changes in AVR by households in pure control villages at midline and endline relative to at baseline by estimating the following specification:

$$y_{itv} = \beta_0 + \beta_1 ML_t + \beta_2 EL_t + X_{iv}^{'}\beta_3 + \gamma_v + \gamma' D_v + \theta' A_v + \epsilon_{itv} \qquad (5)$$

where y_{itv} is the value of the outcome of interest for household i at time t in village v; ML_t and EL_t are binary variables that equal one for data collected during the midline and endline survey rounds, respectively, and zero otherwise; and γ_v represents a village fixed-effect.

Improved agricultural productivity and food security

1.1.4. Does training on the private benefits of CR induce compost production by treated households? Compared to households with only the public health information treatment, i.e., does the content of the training matter, or might inducing CR prompt composting even without compost-related messaging?
We will estimate the specification outlined in equations (2) and (3) and check for significant differences between the estimated coefficients representing the binary variables for villages assigned to treatment arms A, B and C and those between local controls and treated households.

1.1.5. Does training on the private benefits of compost from CR spill over to non-treated neighbors (i.e., local controls) to induce them to engage in CR and compost production?
We will limit the analytical sample to households in villages assigned to treatment arms B and C (which will receive information on private benefits) and the pure control arm, and estimate equation (3).
Does that effect emerge in villages with both private and public health benefits information treatments?
We will estimate the specification shown in equation (4) using the full sample of households and check for significant differences between the estimated coefficients representing the binary variable for local controls and treated households within each treatment arm (A, B and C).

1.1.6. Does training on the private benefits of compost from CR cause increased agricultural total factor productivity (value of total output divided by value of all inputs) and profitability? Does that effect emerge in villages with only public health benefits information treatments?
We will estimate the specification outlined in equation (2) and check for significant differences between the estimated coefficients representing the binary variables for villages assigned to treatment arms A, B and C. We will also test whether local controls in private benefits treatment villages exhibit comparable gains to households that get the private benefits treatment, using equation (4).
Are those effects greatest for poorer households, who are ex ante less likely to invest in chemical fertilizers and other improved inputs?
We will conduct heterogeneity analyses by wealth. Specifically, we will generate an asset index based on baseline asset ownership, designate above- and below-median households in terms of that index using a binary variable, and estimate equation (2) after including that binary variable as a fully interacted covariate.

1.1.7. <u>Does training on the private benefits of CR and its use in compost production boost food security (as reflected in reduced self-reported months of *soudure* and a reduced coping strategies index)?</u>
We will estimate the specification outlined in equation (2) and check for significant differences between the estimated coefficients representing the binary variables for villages assigned to treatment arms A, B and C. We will also conduct heterogeneity analyses by wealth based on a baseline asset index, as above.

<u>Reduced schistosomiasis</u>

1.1.8. <u>Does training in a village reduce the prevalence of schistosomiasis infection (from self-reported condition and symptoms, as well as from urine and stool sample testing among school children), as we compare treatment village households with pure control households?</u>
For self-reported conditions and symptoms, we will estimate the specification outlined in equation (1). For outcomes relating to urine- and stool-sample testing among children, we will estimate the following two-way fixed-effects (TWFE) specification to account for child-specific unobservables:

$$y_{itv} = \beta_1(ML_t \times T_v) + \beta_2(EL_t \times T_v) + X_{itv}^{'}\beta_3 + \beta_4 y_{iv}^{*} + \gamma_i + \gamma_t + \epsilon_{itv} \quad (6)$$

where y_{itv} is the value of the outcome of interest for child i at time t in village v, which will be a binary indicator variable (=1 if infected, =0 otherwise) to study infection at the extensive margin and a continuous measure of schistosoma egg count to capture infection (severity) at the intensive margin; ML_t and EL_t are binary variables that equal one for data collected during the midline and endline survey rounds, respectively, and zero otherwise; T_v is a binary variable that equals one if child i lives in a village assigned to one of the treatment arms, and zero otherwise; and γ_i and γ_t represent a child- and survey round-specific fixed-effects. We will also estimate this using a panel data censored dependent variable estimator (e.g., CLAD).

<u>Does being in a village trained on the private benefits of CR yield greater reduction in schistosomiasis than being trained on the public health benefits only, presumably because of reduced free riding?</u>
For self-reported conditions and symptoms, we will estimate the specification outlined in equation (2). For outcomes relating to urine- and stool-sample testing among children, we will estimate the following modified version of the ANCOVA specification outlined above:

$$y_{itv} = \beta_1(ML_t \times T_A) + \beta_2(ML_t \times T_B) + \beta_3(ML_t \times T_C) + \beta_4(EL_t \times T_A) + \beta_5(EL_t \times T_B) + \beta_6(EL_t \times T_C) + X_{itv}^{'}\beta_8 + \beta_9 y_{iv}^{*} + \gamma_i + \gamma_t + \epsilon_{itv} \quad (7)$$

where T_A, T_B and T_C are binary variables that equal one if child i lives in a village assigned to treatment arm A, B or C, respectively, and zero otherwise.
<u>Does being trained oneself reduce the prevalence of schistosomiasis, as we compare treatment participants versus local controls?</u>

We will estimate the specification outlined in equation (3). Note that this analysis will only apply to self-reported data on conditions and symptoms.

1.1.9. Does training in a village reduce the severity of schistosomiasis infection conditional on infection (from urine and stool sample testing among school children), as we compare treatment village households with pure control households?
We will estimate the TWFE specification outlined in equation (6).
Does being in a village trained on the private benefits of CR yield greater reduction in schistosomiasis egg loads than being trained on the public health benefits only, presumably because of reduced free riding?
We will estimate the TWFE specification outlined in equation (7).

Pro-social behavior and property rights

1.1.10. Does the pre-intervention level of prosociality predict an individual's contribution to AVR? We test whether higher endline contributions in the standard donation game are associated with higher contributions to AVR as measured from the household survey for households with knowledge on public health benefits, according to the following regression specification:

$$y_{iv} = \beta_0 + \beta_1 C_{iv} + X'_{iv}\beta_2 + Z'_{iv}\beta_3 + \gamma' D_v + \theta' A_v + \epsilon_{iv} \qquad (8)$$

where C_{iv} is the standard donation game contribution for household i in village v, Z'_{iv} are controls for the village's treatment arm, and the other variables are defined as before. As a robustness check, we will also run a specification with village level fixed-effects instead of village level controls.
Furthermore, to specifically test whether prosocial households respond more to public health benefits information, we will alter specification (8) as follows:

$$y_{iv} = \beta_0 + \beta_1 C_{iv} + \beta_2 C_{iv} T_{A,C,end} + X'_{itv}\beta_3 + Z'_{iv}\beta_4 + \gamma' D_v + \theta' A_v + \epsilon_{iv} \qquad (9)$$

where $T_{A,C,end}$ is a binary variable that is 1 if the household is part of a village in treatment arms 1 or 3 and the time is endline, and 0 otherwise. According to the hypothesis, we should find that β_2 is positive and significant.

Do the information interventions affect contributions in the donation game? Do such effects spill over from treated households to local controls?
We will use specifications according to equations (2) and (3), with the individual's contribution to the standard donation game as outcome variable.

How does an individual's propensity to donate relate to the individual's and the community's observable characteristics?

Based on the baseline data and the following specification, we test how individual and village characteristics, in particular Lockean beliefs, affect contributions in the standard and impure donation game:

$$y_{iv} = \beta_0 + X'_{itv}\beta_1 + Z'_{iv}\beta_2 + B'_{iv}\beta_3 + \epsilon_{iv} \qquad (10)$$

where B'_{iv} is a battery of variables from the household survey beliefs module, and all other variables are as previously defined.

1.1.11. Does promoting the private benefits of a common pool resource (aquatic vegetation) induce a change in beliefs about property rights? Compare private benefits arms to public health-only arm and pure control arm using beliefs module of household survey. Supplement with qualitative insights from focus group discussions.

1.1.12. Does training in a village change children's school participation (as observed at the school for primary-school-aged children present in study households at baseline)? We will estimate the specification outlined in equation (1). We will also test for within-village spillovers from treated households to local control households by estimating the specification outlined in equation (3). The main outcome is observed school participation: an indicator equal to one if a child is enrolled in school and observed attending school by the study team, and zero otherwise. We will then break down this outcome into observed school enrollment as a measure of school participation on the extensive margin and, conditional on enrollment, observed school attendance as a measure of school participation on the intensive margin. In addition to regressing on outcomes from midline and endline, we will also run regressions on school attendance from "post-midline" data collected in the rainy season near the end of the 2023-2024 school year. Since these outcomes were not collected at baseline, we will use the value reported by households in the baseline household survey as a proxy for the baseline value of the outcome of interest.

2. Secondary outcomes

2.1. Household level

2.1.1. Does training in a village reduce individuals' number of days of work or school lost due to ill health (from self-reported conditions and symptoms)?
We will estimate the specification outlined in equation (1) for each of these two outcomes. We will also test for within-village spillovers from treated households to local control households by estimating the specification outlined in equation (3).

2.1.2. Does training in a village change children's school participation and educational attainment (from self-reported measures on school-aged individuals)?
We will estimate the specification outlined in equation (1). We will also test for within-village spillovers from treated households to local control households by estimating the specification outlined in equation

(3). Outcomes include highest completed grade level as a measure of educational attainment, current school enrollment as a measure of school participation on the extensive margin, and self-reported school attendance as a measure of school participation on the intensive margin.

2.1.3. Do individuals change their contributions when a pure public good is turned into an impure public good? We will use the following regression equation to examine whether individuals contribute more or less in the impure donation game compared to the standard donation game using the following regression equation:

$$y_{ikv} = \beta_{i0} + \beta_1 I_{ikv} + \beta_2 I_{ikv}\delta_i + \beta_3 O_{ikv} + \epsilon_{ikv} \qquad (11)$$

where k is a subscript that indexes the type of game played, y_{ikv} is the outcome for individual i in village v and for game k, I_{ikv} is a binary variable that is 1 if the observation is from the impure donation game and zero otherwise, O_{ikv} is a binary variable that is 1 if the impure game was played before the standard game and zero otherwise, δ_i is a binary variable that is 1 if the individual in the standard donation game contributed more than the threshold value (CFA 200) and zero otherwise, and $\beta_{i0}$is an individual fixed-effect. $\beta_2$will be negative if private benefits result in crowding out community motivations, and will be positive if the existence of private benefits results in a more positive attitude towards public contributions. We will complement this with an alternative version where individual fixed-effects are replaced with a battery of controls at both village and individual level for robustness (see equation 8) .

2.2. Water access point-level

2.2.1. Changes in water use patterns from water point monitoring data. For questions at water access point or community scale, we have far fewer degrees of freedom. We will use regression specifications generally of the form:

$$y_{jv} = \beta_0 + \beta_1 T_v + \beta_3 y^*_{jv} + \gamma' D_v + \Theta' A_v + \psi M_{jv} + \epsilon_{jv} \qquad (12)$$

where y_v is the outcome of interest for water access point j in village v at midline or endline; T is a vector of binary variables that equal one if the village is randomly assigned to one of the three treatment arms, and zero otherwise; D_v is the four element vector of distance (in minutes walking to the nearest village in each of the four experimental arms, with a zero indicating the village is in that treatment arm); M is a binary indicator variable taking value one for water access points that are missing from the baseline sample and zero those included in the baseline sample; and y^*_v is the baseline value of the outcome of interest, which is set to zero in the case of water access points added after baseline. Having established baseline balance among communities and water points, we should be able to use the random variation in treatment assignment, with control for baseline conditions and for distance to other treated villages, to identify the effects of our information intervention at village scale.We are especially interested in how information treatments affect snail and aquatic vegetation populations, where snail population counts come from the dipnet sweeps and vegetative cover come from both sweeps and drone imagery.

2.2.2. Changes in water quality. We want to monitor and test for unintended aquatic ecology consequences of the intervention. To do this, we estimate a variant of equation (12), now adding the contemporaneous value from the upstream water control point as a regressor, so as to control for exogenous changes in water quality that affect the system upstream of (and thus unaffected by) the local intervention. More specifically, we estimate the regression

$$y_v = \beta_0 + \beta_1 T_v + \beta_3 y_v^* + \gamma' D_v + \delta y_v^{uc} + \epsilon_v \qquad (13)$$

where y_v^{uc} is the dependent variable value in the same period from the upstream water control point matched to the water access point under study.

2.2.3. In addition to conducting a Before-After-Control-Impact analysis on water quality and aquatic biodiversity in the villages receiving one of the four treatment arms, we have also designed our sampling to compare treatment arm villages to upstream and downstream sites that are not receiving any treatment. The value of this is that we can assess whether our treatments at water access points are influencing downstream villages. Treatments cannot affect upstream villages, which provide a natural control. To test the hypothesis that treatments disrupt downstream aquatic ecology, we will compare the closest upstream and downstream villages to a village receiving a treatment using a paired test with the distance of each upstream and downstream from the treatment arm village as a covariate. Water quality variables and vegetation weight will be analyzed with normal error distributions, whereas organismal counts will be analyzed with either Poisson or negative binomial error distributions (compared with AIC).

2.3. Community scale

2.3.1. Do information treatments induce changes in natural resource tenure of aquatic vegetation and/ or other, unrelated common pool resources? We will use the qualitative data collected during the focus groups and perform content analysis and thematic analysis to analyze the presence and shape of particular concepts, in particular property rights, privatization, and community control.

Baseline balance

We will conduct balance analyses across all primary and secondary outcomes that were measured at baseline. We will also conduct baseline balance analyses for all variables used as controls in the regressions above. Balance analyses will include both t-tests of differences between treated and untreated, as well as F-tests of the joint null that the vector of outcomes and the vector of control variables are statistically equivalent between treated and control. If baseline imbalance is discovered for more than five percent of variables, we will include the unbalanced covariates as additional controls in our analyses.

Missing data

We will assess the rate of missingness for each outcome of interest at midline and endline. If the missingness rate is less than or equal to 20 percent, we will continue with the analyses outlined above. However, if the missingness rate is greater than 20 percent, we will no longer report analyses for that outcome variable.

Following Lin et al. (2016), we will account for missing data on covariates as follows:

- Observations with missing covariate values will be included in the regressions that estimate treatment effects as long as the outcome measure and treatment assignment are non-missing.
- If no more than 10 percent of the covariate's values are missing, we will recode the missing values to the overall sample mean (or, alternatively, the sample median if we observe that the covariate is not symmetrically distributed).
- If more than 10 percent of the covariate's values are missing, we will include a missingness dummy as an additional covariate and recode the missing values to the overall mean (or, alternatively, the overall median if we observe that the covariate is not symmetrically distributed).

Extreme values

We will test the robustness of our results by excluding extreme values by Winsorizing the relevant outcome variables at the 99, 95 and 90 percent levels.

Multiple outcome and multiple hypothesis testing

As shown in the section on Research Questions above, we have organized our research questions within key outcome "families" based on the level at which outcomes are measured (e.g., household/individual level) and their thematic focus (e.g., diffusion of CR practices). Accordingly, to account for multiple outcome and hypothesis testing, we will control the family-wise error rate when performing multiple hypothesis tests within each of these families of outcomes. We will do so by estimating adjusted p-values using the free step-down resampling methodology of Westfall and Young (1993) as operationalized in the -wyoung- command in Stata. These adjusted p-values will be presented as robustness checks for our main results.

Appendix A: Sample village listing and map

Region	Department	Commune	Villages Name (from census)	Village Name (local)
Saint-Louis	PODOR	GUEDE VILLAGE	AGNAM TONGUEL	
Saint-Louis	DAGANA	NDIAYE	AMOURA	
Saint-Louis	DAGANA	DIAMA	ASSY	
Saint-Louis	PODOR	THILLA BOUBACAR	BAKAO	
Saint-Louis	DAGANA	ROSS-BETHIO	BISSETTE I	
Saint-Louis	PODOR	GAE	BOULEYDI	
Saint-Louis	PODOR	GUEDE VILLAGE	DADO	
Saint-Louis	DAGANA	MBANE	DAGANA	
Saint-Louis	PODOR	DODEL	DARA ALAYBE	
Saint-Louis	PODOR	THILLA BOUBACAR	DARA SALAM	DAR SALAM
Saint-Louis	PODOR	THILLA BOUBACAR	DEGUEMBERE	
Saint-Louis	PODOR	GAMADJI SARRE	DEMBE	
Saint-Louis	PODOR	NDIAYENE PENDAO	DIABOBES	
Saint-Louis	DAGANA	ROSSO	DIADIAM I	
Saint-Louis	DAGANA	ROSS-BETHIO	DIADIAM III	
Saint-Louis	DAGANA	NDIAYE	DIAGAMBAL I	
Saint-LOUIS	DAGANA	DIAMA	DIAMA	
Saint-Louis	PODOR	DODEL	DIAMAL	
Saint-Louis	PODOR	NDIAYENE PENDAO	DIAMEL (DIAMEL DJIERY)	DIAMEL DJIERY
Louga	LOUGA	KEUR MOMAR SARR	DIAMINAR	DIAMINAR KEUR KANE
Louga	LOUGA	KEUR MOMAR SARR	DIAMINAR LOYENE	
Saint-Louis	PODOR	GAMADJI SARRE	DIARRA	
Saint-Louis	DAGANA	RONKH	DIAWAR	

Saint-Louis	PODOR	GUEDE VILLAGE	DIEGUESS DAROU SALAM	GUEDE VILLAGE
Saint-Louis	DAGANA	DIAMA	DIOSS PEULH	PEULH DIOSS
Saint-Louis	PODOR	GAMADJI SARE	DIOUDE	
Saint-Louis	SAINT-LOUIS	RAO	DIOUGOP	
Saint-Louis	PODOR	DODEL	DODEL	
Saint-Louis	PODOR	GAMADJI SARE	DODEL	DARA ALAYBE
Saint-Louis	PODOR	GUEDE VILLAGE	DONAYE	
Saint-Louis	PODOR	GUEDE VILLAGE	DOUE	
Saint-Louis	DAGANA	ROSS-BETHIO	EL DEBIYAYE MARAYE II	MARAYE
Saint-Louis	DAGANA	DIAMA	EL MOHAMED AMAR	EL MOHAMED LAMAR
Saint-Louis	PODOR	FANAYE	FANAYE DIERY	
Saint-Louis	PODOR	FANAYE	FANAYE WALO	
Louga	LOUGA	KEUR MOMAR SARR	FĖTO	
Saint-Louis	PODOR	GUEDE VILLAGE	FONDE ASS	
Saint-Louis	DAGANA	MBANE	FOSS	
Saint-Louis	PODOR	GAMADJI SARRE	GAMADJI SARRE	
Louga	LOUGA	KEUR MOMAR SARR	GANKETTE BALLA	
Louga	LOUGA	KEUR MOMAR SARR	GAYA	
Saint-Louis	DAGANA	NDIAYE	GNITH	
Saint-Louis	DAGANA	NDIAYE	GOBAK	
Saint-Louis	PODOR	GUEDE	GUEDE	BIRGAL (neighborhood in Guede)
Louga	LOUGA	KEUR MOMAR SARR	GUEO	
Saint-Louis	DAGANA	DAGANA	GUEUM YALLA	
Saint-Louis	DAGANA	BOKHOL	GUIDAKHAR	
Saint-Louis	PODOR	GUEDE	H3 PETEL DIEGUESS	DIABBE (neighborhood in Guede)
Saint-Louis	PODOR	NDIAYENE PENDAO	KADIOGUE (DIABOBES II)	KADIOGNE

Saint-Louis	DAGANA	RONKH	KASSACK NORD	
Saint-Louis	DAGANA	DAGANA	KEUR BIRANE KOBAR	
Saint-Louis	DAGANA	BOKHOL	KHARE	
Saint-Louis	DAGANA	RONKH	KHEUNE	
Saint-Louis	DAGANA	RONKH	KHOR	
Saint-Louis	PODOR	GUEDE VILLAGE	KODITH	
Saint-Louis	PODOR	GUEDE VILLAGE	LERABE	
Saint-Louis	DAGANA	MBANE	LEWAH (TEMEYE LEWAH)	LEWA (TEMEYE LEWA)
Saint-Louis	PODOR	NDIAYENE PENDAO	LOBBOUDOU DOUE	
Saint-Louis	DAGANA	ROSS-BETHIO	MALLA	
Saint-Louis	DAGANA	MBANE	MALLA TACK	
Saint-Louis	DAGANA	RONKH	MBAGAME	
Saint-Louis	DAGANA	NDIAYE	MBAKHANA	
Saint-Louis	PODOR	PODOR	MBANTOU	
Saint-Louis	DAGANA	NDIAYE	MBARIGO	
Saint-Louis	DAGANA	DIAMA	MBERAYE	
Saint-Louis	DAGANA	NDIAYE	MBEURBEUF	
Saint-Louis	DAGANA	DAGANA	MBILOR	
Saint-Louis	DAGANA	NDIAYE	MBOLTOGNE	CROISEMENT SAVOIGNE
Saint-Louis	DAGANA	DIAMA	MBOUBENE PEULH	MBOUBENE NARR
Saint-Louis	PODOR	GUEDE VILLAGE	MBOYO	
Louga	LOUGA	KEUR MOMAR SARR	MERINA GEWEL	
Saint-Louis	Dagana	NDIAYE	MINGUENE BOYE	
Saint-Louis	DAGANA	RONKH	NADIEL I	NADIEL
Saint-Louis	DAGANA	ROSS-BETHIO	NAERE	
Saint-Louis	Dagana	NDIAYE	NDELLE BOYE	

Saint-Louis	DAGANA	ROSS-BETHIO	NDER	
Saint-Louis	DAGANA	MBANE	NDIAKHAYE	
Saint-Louis	SAINT LOUIS	GANDON	NDIALAKHAR WOLOF	NDIALAKHAR WOLOF
Saint-Louis	DAGANA	ROSS-BETHIO	NDIAMAR	SOULOUL
Saint-Louis	PODOR	GUEDE VILLAGE	NDIAWARA	
Saint-Louis	SAINT-LOUIS	RAO	NDIAWDOUNE	
Saint-Louis	DAGANA	ROSS-BETHIO	NDIAYE MBERESSE (NDIAYE NGAINTHE)	KARAMATOU
Saint-Louis	PODOR	THILLA BOUBACAR	NDIAYENE PENDAO	NDIAYENE SARE
Saint-Louis	PODOR	NDIAYENE PENDAO	NDIAYENE SARE	NDIAYENE PENDAO
Louga	LOUGA	KEUR MOMAR SARR	NDIBE	
Saint-Louis	DAGANA	RONKH	NDIETENE	
Saint-Louis	DAGANA	NDIAYE	NDIOL MAURE	
Saint-Louis	DAGANA	NDIAYE	NDIOUNG MBERESSE	NDIOUGUE MBERESSE
Saint-Louis	DAGANA	NDOMBO	NDOMBO	NDOMBO SANDJIRI
Saint-Louis	DAGANA	NDOMBO	NDOMBO ALARBA	
Saint-Louis	PODOR	DODEL	NDORMBOSS	NORMBOSS
Saint-Louis	PODOR	GUEDE VILLAGE	NGAOULE	
Saint-Louis	SAINT LOUIS	GANDON	NGAYE	
Saint-Louis	PODOR	NDIAYENE PENDAO	NGEUNDAR (GARAGE NGUENDAR)	NGEUNDAR
Saint-Louis	DAGANA	NDIAYE	NGOMENE	
Saint-Louis	PODOR	GUEDE VILLAGE	OURO MADIHOU	
Saint-Louis	DAGANA	ROSS-BETHIO	PAKH	
Saint-Louis	PODOR	DODEL	PATHE GALLO	
Saint-Louis	DAGANA	RONKH	RONKH	

Saint-Louis	DAGANA	ROSS BETHIO (ODABE NAWAR)	ROSS BETHIO (ODABE NAWAR)	ODABE NAWAR
Saint-Louis	DAGANA	MBANE	SANEINTE TACQUE	SANEINTE
Saint-Louis	DAGANA	NDIAYE	SAVOIGNE PEULH	KEUR SAMBA DIAM
Saint-Louis	DAGANA	DIAMA	SAVOIGNE PIONNIERS	SAVOIGNE PIONIERS
Saint-Louis	DAGANA	MBANE	SYER	
Saint-Louis	DAGANA	DIAMA	TABA TREICH	
Saint-Louis	DAGANA	MBANE	TEMEYE	TEMEYE THIAGO
Saint-Louis	DAGANA	RONKH	THIAGAR	
Saint-Louis	PODOR	THILLA BOUBACAR	THIANGAYE	
Saint-Louis	PODOR	GAMADJI SARRE	THIELAO	THIELLAO
Saint-Louis	PODOR	NDIAYENE PENDAO	THIEWLE	
Saint-Louis	DAGANA	NDIAYE	THILENE	
Saint-Louis	Dagana	NDIAYE	THILLA	
Saint-Louis	DAGANA	ROSSO	TIGUETTE	
Saint-Louis	DAGANA	NDIAYE	TREICH PEULH	
Saint-Louis	DAGANA	ROSS-BETHIO	YAMANE	
Saint-Louis	DAGANA	RONKH	YETTI YONI (BOUNTOU NDIEUGNE)	YETTI YONE

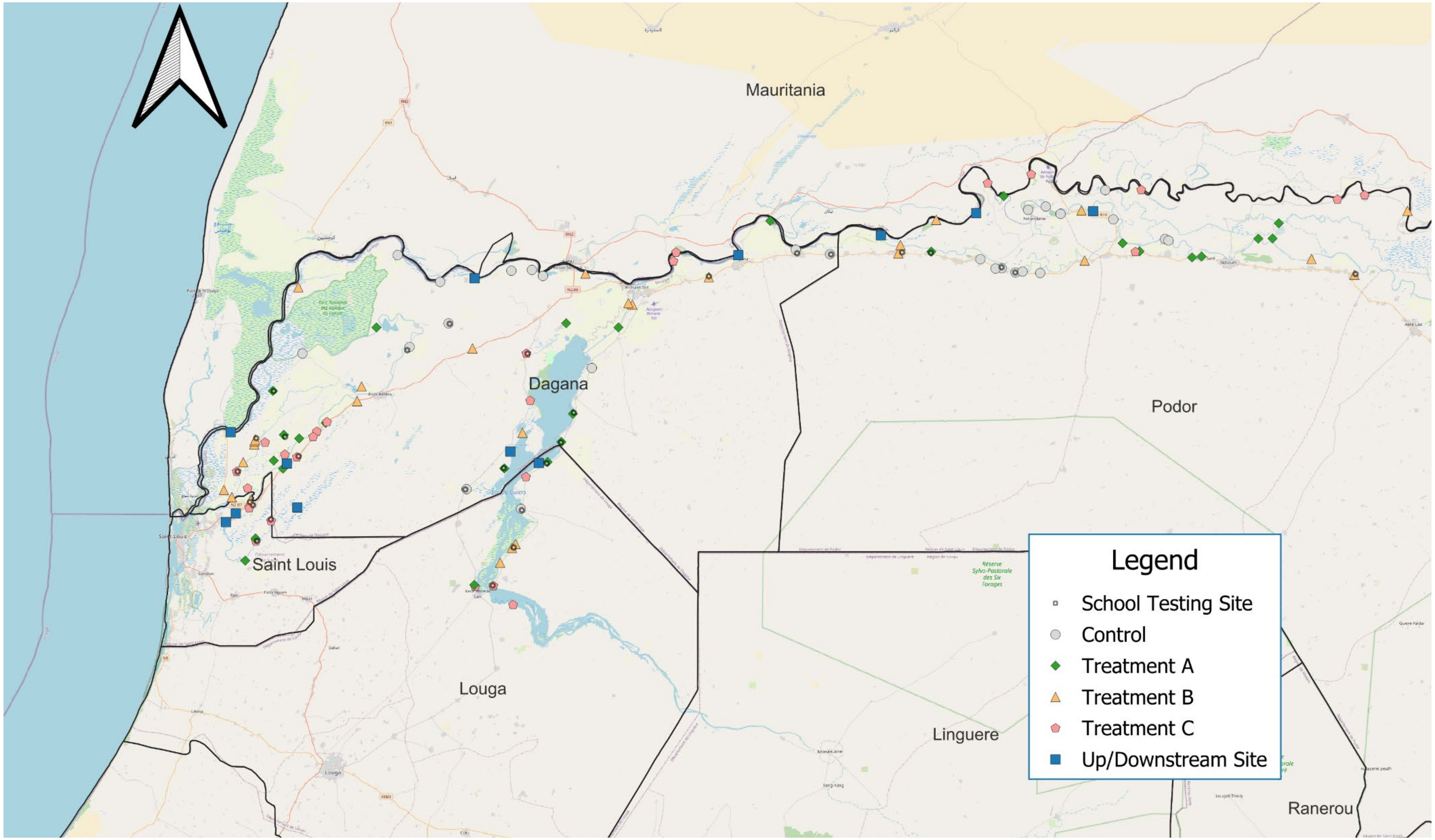


Figure A1: Map of area of the Senegal river and the lac de Guiers showing the location of the study villages. The "C" in the middle of the symbol denotes villages with human parasitological testing. Note: the map has been updated from that in the original PAP to correct errors in a couple of schools' geocoordinates.

Appendix B: Drone imagery data collection and analysis protocol

Imagery of the full water-access point will be captured via a Micasense RedEdge-MX multispectral camera attached to a DJI Inspire 2 drone. The Micasense RedEdge-MX camera maintains 5 sensors, each dedicated to a specific portion of the electromagnetic spectrum: Blue (475 nm center, 32 nm bandwidth), Green (560 nm center, 27 nm bandwidth), Red (668 nm center, 14 nm bandwidth), Rededge (717 nm center, 12 nm bandwidth), and Near-infrared (842 nm center, 57 nm bandwidth). Calibration information will be collected with an associated down-welling light sensor which will account for changes in cloud coverage or light intensity throughout the drone flights in addition to an image of a calibrated reflectance panel.

The flight altitude is 100 meters. The distance covered extends from left to right at approximately 150 meters and follows the direction of the wind to avoid excessive battery consumption. All frequented water points in a village are flown over with the drone. The water points are approximately 50 to 500 meters apart. Drone overflight is authorized under the number 005871/MINT/DGPN/DST/DAM from May 12, 2022 by the Ministry of the Interior for a period of 12 months recently renewed under the number 011936/MINTSP/DGPN/DST/DAM from November 14, 2024 and also valid for 12 months.

After image collection, an object-based image analysis (OBIA) workflow will be utilized for pre-processing imagery before running a machine learning model for *Ceratophyllum* identification (Chabot et al. 2018) (Chabot et al., 2018). An OBIA has been selected as it is well suited to explore the heterogeneity of wetlands and aquatic systems (Dronova, 2015; Chabot et al., 2016; Husson et al., 2016; Chabot et al., 2018; Visser et al., 2018). Imagery will be radiometrically calibrated and stitched before images are mosaiced and rendered into absolute reflectance maps (pixel values ranging from 0-1). Multiple segmentation along spectral characteristics will be implemented—allowing for discrimination between submerged and floating aquatic vegetation (Chabot et al., 2018). The performance of the trained machine learning classifier will be evaluated using the classified, drone-acquired imagery. Random forest was chosen due to its suitability in high-dimensional feature spaces and accounting for overfitting (Pal, 2005). False positives will be classified as instances where an object is labeled as a particular class but does not actually belong to that classification. False negatives will be classified as instances when an object is not labeled with the appropriate classification by the model. The accuracy of the model on the imagery classification will be determined through kappa, AUC, precision, recall, and F1 score. The amount of *Ceratophyllum* present per water access point will be determined as a proportional coverage.

Appendix C: Household and Community Surveys (including consent and focus group discussion scripts) and Post-Training Comprehension Questionnaire