The Economic Impact of Schistosomiasis

Daniele Rinaldo[1*]

Javier Perez-Saez[2†]

Penelope Vounatsou[3‡]

Jean-Louis Arcand[1§]

[1] The Graduate Institute of International and Development Studies, Geneva
[2] Johns Hopkins University, Bloomberg School of Public Health, Baltimore
[3] Swiss Tropical and Public Health Institute, Basel

Abstract

Schistosomiasis is a chronic and debilitating neglected tropical disease that is endemic in many developing countries. The disease is contracted by direct contact with waters infested by the pathogen, whose life cycle requires an intermediate host (a freshwater snail) whose habitats are broadly expanded by water resources management projects. We study the impact of schistosomiasis on economic development by means of its effect on agricultural production. We focus on Burkina Faso, a sub-Saharan African country where the disease is endemic and agriculture is mostly of the subsistence variety, in order to establish a baseline for the worst possible burden. We estimate this impact to be large, negative and consistent with the disease being a poverty-reinforcing productivity shock, and establish causality of the effect by instrumenting the disease intensity with the density of the snail hosts. We introduce two methodological innovations: econometrics based on high-resolution prevalence maps and the use of machine learning techniques to identify the disease burden. We present evidence of the feedback between disease diffusion and development: the creation of large dams, while boosting agricultural production, magnifies the adverse effects of the disease.

*daniele.rinaldo@graduateinstitute.ch
†jperez5@jhu.edu
‡penelope.vounatsou@swisstph.ch
§jean-louis.arcand@graduateinstitute.ch

Work leading to this study was funded by the Swiss Network of International Studies. The paper was conceived by D.R. and J.L.A. The paper was written by D.R., J.P.S. and J.L.A. Modeling, identification strategy and estimations done by D.R. Geolocalization of the villages, mapping and intensity measures done by J.P.S. The two disease prevalence maps were provided by P.V. The authors are greatly indebted to Théophile Mande (mandetheophile@gmail.com) in Ouagadougou for his help in obtaining the survey datasets. We also thank the Schistosomiasis Control Unit of the Ministry of Health of Burkina Faso for access to parasitological data, and the Department of Sectoral Statistics at the Ministry of Agriculture and Hydrological Development of Burkina Faso for their collaboration in allowing access to the agricultural datasets. The authors declare no further competing interests.

1 Significance statement

We study the economic impact of schistosomiasis, a parasitic disease with a complex transmission cycle which is endemic in many developing countries, by means of its effect on agricultural production. We create a novel dataset that combines high-resolution disease prevalence maps with detailed agricultural and household surveys. We find a large, negative and nonlinear causal effect of the disease on yields. Our empirical results uncover a negative feedback loop between disease dynamics and water resources development, and are consistent with schistosomiasis constituting a poverty-reinforcing productivity shock.

2 Introduction

Schistosomiasis (*Bilharzia*) is a debilitating neglected tropical disease that affects an estimated 140 million people, more than 85% of whom live in sub-Saharan Africa (James et al., 2018). Claiming 3 million disability-adjusted life years per year in the past decade, its global burden ranks second only to malaria among parasitic diseases (James et al., 2018). Severe morbidity due to schistosomiasis results from the accumulation of eggs laid by flatworms of genus *Schistosoma* in the tissues of the human host, leading to a chronic inflammatory response (Colley et al., 2014). The parasite species that cause the two main forms of the disease (intestinal and uro-genital schistosomiasis) present a complex life cycle involving two reproduction phases, the first asexual in specific species of aquatic snail intermediate hosts, followed by sexual in the human host. Infection occurs through skin penetration by water-motile schistosome larvae, which, once matured, mate and secrete eggs that, exiting the human host through urine or feces, perpetuate the parasite's lifecycle. A large fraction of these eggs remains trapped in the tissues surrounding the bladder or the intestine, eliciting the chronic inflammation that constitutes the root of schistosomiasis-induced morbidity (Colley et al., 2014). When untreated, advanced forms of schistosomiasis lead to kidney failure, bladder cancer, liver fibrosis (Richter, 2003; Andrade et al., 2017), as well as heightened risk of HIV transmission (Mbabazi et al., 2011). The highest parasite burden is usually borne by children and the disease has been linked to anemia, stunting, cognitive deficits leading to poor school performance and higher drop-out rates (King, 2010; Ezeamama et al., 2018). Due to these life-long impacts, schistosomiasis exerts large health, social and financial burdens on infected individuals and households (Adenowo et al., 2015), and is considered to be both a consequence and a cause of poverty (King, 2010).

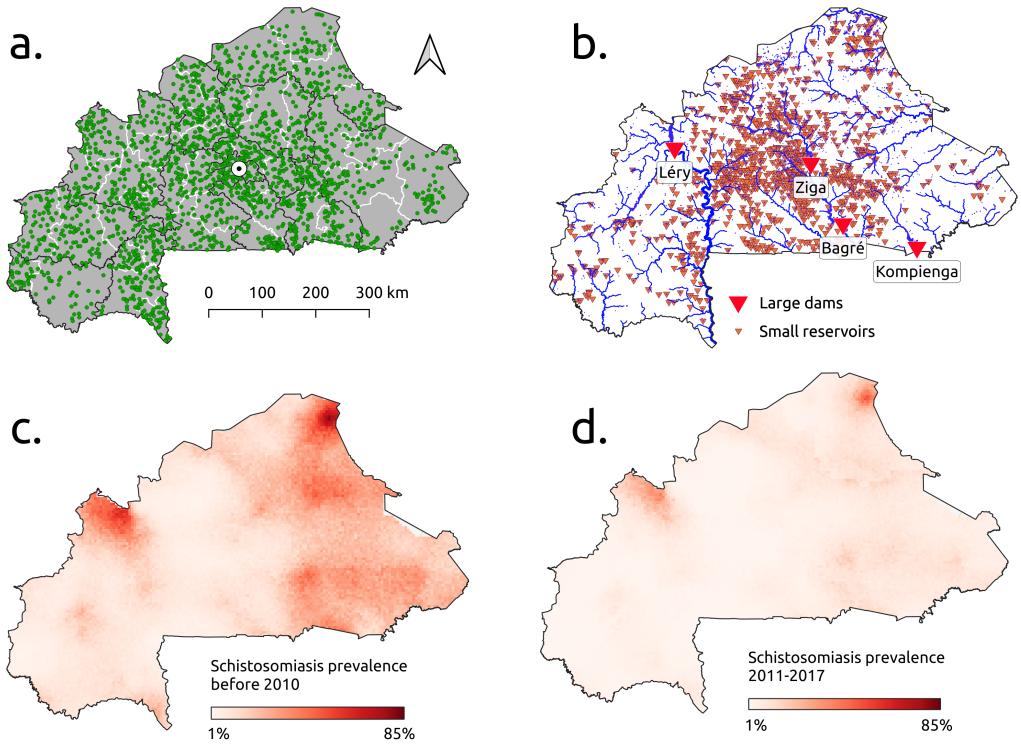
Water resources development aimed at alleviating poverty in rural schistosomiasis-endemic communities in sub-Saharan Africa has been shown to exacerbate disease transmission (Steinmann et al., 2006). Dams and irrigation schemes increase the abundance of the aquatic snails that serve as the intermediate hosts of schistosomes, and also increase the frequency and density of human-water contacts during which infection can occur (Boelee et al., 2009; Perez-Saez et al., 2015; Diakité et al., 2017). This effect is particularly marked in water-constrained regions, possibly due to the concentration of human-water contacts in the few available water points (Steinmann et al., 2006; Perez-Saez et al., 2017). In developing countries, the sector which one would naturally expect to be the most affected by the disease is agriculture, particularly in its subsistence form. This is because populations that rely heavily on agricultural production are the ones that are the most exposed to infection and ultimately suffer the highest disease burden. While schistosomiasis may be an example of a mechanism leading to a 'poverty trap', its economic impact and the associated tradeoffs between water resources development and public health concerns have not been rigorously measured.

General results concerning the relationship between diseases and economic development can be found in Acemoglu and Johnson (2007), Acemoglu et al. (2003), Bleakley and Lange (2009) and Audibert (2010), with the latter focusing on the effect of endemic diseases on agricultural productivity. Despite extensive evidence concerning the long-term health effects of various diseases, there have been few attempts to quantify their economic impact. This is particularly true of parasitic diseases such as schistosomiasis. St. Lucia was one of the first countries studied in an effort to ascertain the impact of schistosomiasis on labour productivity (Baldwin and Weisbrod (1974); Weisbrod et al. (1974)): very little was detected, but mismeasured variables and lack of data hindered the effort. The effect of schisosomiasis on agricultural production was first studied by Foster et al. (1967): again, only small negative effects were found. The quasi-experiment carried out by Audibert and Etard (1998) in Mali found no direct effects on rice production, but did find effects on the use of labour and other resources within households. Conversely, Audibert (1986) found a negative effect of schistosomiasis on rice production for infected households in the Cameroon. In Miguel and Kremer (2004), the authors evaluate a deworming project that includes schistosomiasis and establish the former's large health benefits, although without evidence of treatment effects on academic performance.

We focus on Burkina Faso, a country where schistosomiasis is endemic (Poda et al., 2004), in order to establish a baseline case for what is potentially a large negative effect, and where there are less confounders generated by the interdependence of economic sectors found in more diversified economies. Burkina Faso is a low-income, landlocked Sub-Saharan country with very limited resources. In 2016, 80% of its 18 million inhabitants worked in the agricultural sector. Burkinabé agriculture is of the subsistence variety, with low crop and livestock productivity (World Bank Results, 2017). Diversification in the sector is low, although increasing, with cotton being the most important cash crop. Large and small-scale water resource development projects have been completed in the past 30 years to support agricultural activities and reduce climate vulnerability (Fig. 1b) (Cecchi et al., 2007), which have however exacerbated the prevalence of malaria and schistosomiasis (Poda et al. (2004); Boelee et al. (2009)). Historically Burkina Faso has experienced high prevalence of both intestinal and uro-genital schistosomiasis (Fig. 1c) (Poda et al., 2004). Mass chemotherapy campaigns initiated by the Schistosomiasis Control Initiative since 2005 have been successful in most regions with a national mean prevalence of around 5% in school age children in 2010 (Fig. 1d) (Ouedraogo et al., 2016).

Because of the complexity of the dynamics of schistosomiasis and its interlinkages with a large set of socioeconomic variables, identifying the relationship between the disease and economic development via its effect on productivity requires the use of detailed agricultural and household datasets, as well as precise information on disease prevalence on a large spatial scale. Recent developments in disease mapping allow us to obtain high resolution prevalence maps which have been used in a variety of public health contexts (Lai et al., 2015) but have not hitherto been paired with household and plot-level data. We first create a large dataset for Burkina Faso that merges household and plot-level information with village-level disease prevalence data. We then proceed to estimate the effect of schistosomiasis on agricultural production with a variety of econometric techniques. We contribute to the literature by the novel use of high-resolution disease maps coupled with machine learning methods in the estimation of an agricultural production function. Furthermore, we address both endogeneity and measurement error by exploiting natural heteroskedasticity as an internal instrument. In order to address endogeneity and establish causality of the effect of schistosomiasis, we use the density of the aquatic snails that serve as intermediate host of the disease as an instrumental variable (henceforth, IV).

Figure 1 Overview. (a) EPA Villages, the capital Ouagadougou (white point) and level 1 (regions, black lines) and level 2 (provinces, white lines) administrative subdivisions. (b) River network (blue lines, width proportional to upstream area) and water resources infrastructure in the country. (c) Estimated schistosomiasis prevalence up to 2010 (Lai et al., 2015). (d) Estimated schistosomiasis prevalence for 2011-2017.



3 Data

The agricultural and household survey data were obtained from the National Institute of Statistics and Demography (INSD) in Ouagadougou. The annual plot-level agricultural dataset provides detailed information on crops, yields, inputs and pesticides, plot characteristics and labour type for the 2003-2017 period for ca. 200,000 unique plots cultivated by roughly 20,000 households in 1,950 villages (details are provided in the Appendix). Surveyed villages were distributed relatively uniformly across the country (Fig. 1a). The household survey data cover the 1996-2017 period and include detailed information on household characteristics and demographics, as well as livestock type and use. We first merge and synchronize both survey datasets, and geolocalize the villages. We include a large variety of climatic remote-sensing data including precipitation amounts as well as derived statistics, temperature, and vegetation indices. Full information on the dataset, summary statistics and details concerning the covariates used in the analysis can be found in the Appendix .

We then merge the dataset with two high-resolution maps of the prevalence of schistosomiasis in school-aged children that give an estimate of the predicted prevalence at a pixel resolution of 5×5 kilometers. The first map applies up until 2010 (Fig. 1c), the second thereafter (Fig. 1d). The estimated prevalence is a joint measure of both *S. haematobium* and *S. mansoni*, and is obtained by means of Bayesian geostatistical analysis. For details concerning the prevalence map, we refer to Lai et al. (2015), which presents the first of the two maps in full detail. Because of resolution of the disease maps, prevalence is constant within each village: households belonging to the same village will be assigned the same level of schistosomiasis prevalence. For villages located at pixel borders, we take the average of the measures on each side. We then translate estimated prevalence into a measure of schistosomiasis infection intensity in terms of the average number schistosome egg-output per person. This was done by assuming a negative binomial distribution of egg counts in urine and stools fit to parasitological data (see Appendix for details). Moreover, the health consequences of schistosomiasis are more directly linked to infection intensity (measured bgy egg-output) than to prevalence (Audibert (1986); King and Dangerfield-Cha (2008)). The malacological data were collected in two field sites located along the South-North climatic gradient between the Sudanian and Sahelian climatic regions. A detailed description can be found in Perez-Saez et al. (2016).

4 Model and estimation methods

4.1 Estimation at the plot level

In order to estimate the productivity effects of schistosomiasis at the plot level we begin by specifying the agricultural production technology as:

$$Y_{ihjt} = A_{hj}\phi_{hjt}^{-\theta}F(X_{ihjt})e_{ihjt}, \tag{1}$$

where Y_{ihjt} is the yield (output per hectare) of plot i, farmed by household h in village j at time t, ϕ_{hjt}, which is common to all plots cultivated by a given household, represents the direct effect of schistosomiasis on productivity, calibrated by a parameter θ, A_{jh} are household time-invariant productivity shifters, some of which may be unobservable, $F(.)$ is the production function and X_{ihjt} is the matrix of total inputs. Lastly, $e_{ihjt} = \exp(\alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt})$ represents plot-level unobservables, which we decompose into crop- (c), household- (h) and time-specific components (t). Since the "real" effect ϕ of the disease is unobservable, we approximate it by the measure of infection intensity in terms of mean egg-output per person discussed earlier, denoted by I_{jt}. Taking logarithms then results in the quasilinear model:

$$y_{ihjt} = \tilde{A}_{jh} - \theta\ I_{jt} + \tilde{F}(X_{ihjt}) + \alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt}, \tag{2}$$

where our goal is to identify the parameter θ. We begin by imposing a Cobb-Douglas functional form on $F(.)$, which yields a fully linear model. The instrumented model takes the same form as (2). We subsequently estimate the median effect of intensity by means of the quantile panel procedure of Koenker (2004). Note, if the household is assumed to maximize profits over its set cultivated plots, that $\tilde{F}(X_{ihjt}, K_{ihjt})$ must be replaced by $\tilde{F}(X^*_{ihjt}, K_{ihjt})$, where $X^*_{ihjt}(A_{hj}, I_{jt}^{-\theta}, p_{jt}, K_{ihjt}, \epsilon_{ihjt})$ is input use at the optimum, p_{jt} represents input prices and K_{ihjt} represent all other factor of production not under the control of the household (for example, climate). In this case the total marginal effect of schistosomiasis on yield would be given by $-\theta + \sum_{w=1}^{w=W} \frac{\partial \tilde{F}(X^*_w, K)}{\partial X_w} \frac{dX^*_w}{dI}$, where W is the total number of inputs chosen optimally. In the Appendix we show that the specification

given by (1) is compatible with a model in which schistosomiasis affects *effective* labor input and that, empirically, $\frac{dX_w^*}{dI} = 0$ for all $w = 1 \dots W$: schistosomiasis therefore is a pure productivity shock that does not affect optimal input use.

We then relax the restrictive assumptions (such as unitary elasticity of substitution between factors) imposed by the Cobb-Douglas functional form, and use adaptive machine learning methods in order to partial out all confounding and nonlinear effects of the factor inputs X_{ihjt}. The quasi-linear structure of (2), stemming from the identification of schistosomiasis as a productivity shock, allows us to disentangle the effect of the disease from the effect of the other covariates. After obtaining a consistent estimate of the function $\tilde{F}(.)$, we can partial out its predicted values from both log-yield and schistosomiasis infection intensity, thereby identifying the effect of interest, θ. Since determining the precise parametric form of $\tilde{F}$ that best fits the data is of secondary interest in the present context, we use adaptive machine learning methods, which are obscure in terms of interpretation but well-suited for prediction and classification. This procedure, named double/debiased machine learning (DML), was first proposed by Belloni and Chernozhukov (2013) and further expanded in Belloni et al. (2014), Belloni et al. (2017) and Chernozhukov et al. (2018) in order to deal with overfitting due to high-dimensional data and the estimation of treatment effects. Here, we extend this procedure to the regularization of all confounding effects due to the unknown functional form for the production function: this extension is one of the methodological contributions of the paper.[1] The method is readily extended to the case of multiple IVs: each of the IVs is partialled out, and the resulting residual used in a standard IV estimation. The method requires imposing the double Neyman-orthogonal moment conditions:

$$\mathbb{E}[\psi(W, \theta, \eta_0)] = 0, \tag{3}$$

$$\partial_\eta \mathbb{E}[\psi(W, \theta, \eta)]_{\eta=\eta_0} = 0, \tag{4}$$

where $\psi(W, \theta, \eta_0) = ((y - \mathbb{E}[y|X]) - (I - \mathbb{E}[I|X])\theta)(I - \mathbb{E}[I|X])$ is the Neyman-orthogonal score function, $\eta_0 = (\mathbb{E}[y|X], \mathbb{E}[I|X])$, and where X is the complete matrix of covariates; ∂_η is a functional derivative operator, and the second condition imposes validity of the estimators under possible deviations from η_0. For the estimation of the unknown nuisance functions $\mathbb{E}[y|X]$ and $\mathbb{E}[I|X]$, we use random forests, gradient boosting machines and neural networks to obtain their predicted values and identify the parameter of interest θ, and we compare their performance using a mean-squared error criterion.[2] Convergence of the estimator is ensured by cross-validation for out-of-sample fitting. This approach is easily extended to an IV procedure, under the maintained hypothesis that the exclusion restriction for the IVs hold, and adds one stage per instrument to the usual procedure in which the Neyman-orthogonal moment conditions are imposed on each of the IVs. We also estimate a semi-parametric model in order to expose potential non-linearities in the effect of the disease, where θI_{jt} is replaced by the smooth function $f_I(I_{jt})$ in Eq (2).[3]

Our use of the density of snails as an IV requires additional explanation. In Burkina Faso there are two main forms of schistosomiasis, intestinal and uro-genital. The prevalence measure that we use in our analysis is a joint measure, generated by averaging the two forms and subtracting the

[1] The procedure needs to be cross-fit to remove bias. Monte Carlo simulations (available upon request) of different production functions show how non-crossfit DML can actually increase bias.

[2] Note that estimating $\mathbb{E}[y|X]$ is equivalent to estimating $\tilde{F}$. For further references on the estimators, see Hastie et al. (2001).

[3] Here f_I is estimated by minimizing the squared residuals with a smoothing penalty on the second derivative of f_I. The smoothing parameter is chosen by cross-validation, and the degrees of freedom of the interpolating function are chosen iteratively by checking for zero signal on the residuals. The multidimensional splines are fit in a similar manner when studying the interaction of schistosomiasis intensity with other variables (malaria prevalence and distance to dams/rivers).

covariance, as in Lai et al. (2015). We therefore need to instrument the two forms of the disease by including information on the different species of snail hosts. The intestinal form caused by *S. mansoni* is concentrated in the southwest part of the country, and its range is constrained by the presence of its intermediate host snails, *Biomphalaria pfeifferi*, as shown by Perez-Saez et al. (2017). This snail species is not present outside of this region due to its sensitivity to prolonged habitat dryouts which are more common in the Central and Northern parts of the country (Poda et al. (2004); Perez-Saez et al. (2017)). As an instrument for intestinal schistosomiasis we therefore use the mean of *Biomphalaria* abundance for the villages in the southwest. The uro-genital form of the disease, caused by a different species of schistosomes, *S. haematobium*, is spread more uniformly throughout the country due to the ubiquity of the snail species of the genus *Bulinus* which serve as intermediate hosts. The highest prevalence of uro-genital schistosomiasis is found in the northern part of the Sahel. As opposed to *Biomphalaria* spp., *Bulinus* spp. are present in a wide range of natural and man-made habitats throughout the country (Poda et al., 2004), and we use estimates of its abundance in ponds and zones with ephemeral rivers in both the rainy and the dry season in order to capture changes in river size. We use gridded predictions of the seasonal variations of snail densities produced in Perez-Saez et al. (2019). In our dataset, the correlation between disease intensity and snail abundance is positive and strong, especially with *Biomphalaria* in rivers in the southwest, and with *Bulinus* in the northern dry regions where rivers are more ephemeral. The presence of either kind of aquatic snail is directly linked to the presence of both forms of schistosomiasis, whilst having no direct effect on agricultural yields. As such, it constitutes an ideal IV.

4.2 Village aggregation

Aggregating at the village level, the quasilinear model for a representative household is of the form:

$$y_{jt} = \tilde{A}_j - \theta I_{jt} + \tilde{F}(X_{jt}) + \alpha_t + \epsilon_{jt}. \tag{5}$$

The model is then estimated by IV in the same manner as was the case at the plot-level. When controlling for spatial effects and spatial correlation, the model we fit after stacking observations and imposing a Cobb-Douglas functional form is the mixed spatial autoregressive model (SAR) given by:

$$y_{jt} = (1_{2j} - \rho W_j)^{-1}[I_{jt}\theta + \tilde{X}_{jt}\beta + \alpha_t + \epsilon_{jt}], \tag{6}$$

where W_j is the matrix of spatial weights given by the distance between the j villages, ρ is the spatial autoregressive parameter, and $\tilde{X}_{jt}$ is the matrix of factor inputs expressed in logarithmic form. We fit a model where both dependent and independent variables are spatially lagged. In order to check for the presence of spatial random effects, we fit a model with a random term $u(d)$ with mean zero and spatial covariance function $M_\nu(d)$. We adopt the standard choice of a Matérn covariance function. All parameters are then fit by maximum likelihood, allowing us to estimate the spatial correlation between village pairs, and to obtain its rate of decay as distance increases.

5 Results

5.1 Plot- and village-level estimation

We present results for households and villages observed in 2009 and 2011. These years are chosen to maximize the number of observations and provide the best fit and accompanying model diagnostics. Results are broadly similar for other combinations of years for which there are enough

repeated observations. Results are also similar if disease prevalence obtained directly from the maps is used in place of our intensity measure. Additional tables and results are reported in the Appendix. Figure 2 reports the different estimates of θ. Given that the intensity measure is expressed in terms of worm eggs per person, and that its value is in the 0-110 range, the coefficient associated with intensity is expected to be negative and of the order of 10^{-2}-10^{-3}. The point estimate represents the marginal impact of one additional worm egg per person on log yield.

The upper three estimates in Figure 2 impose a Cobb-Douglas functional form, yielding a fully linear model; the first includes year and crop fixed effects, the second adds household ones. The point estimates imply a loss in mean yield of between around 1% and 4%, with villages in the top 5% quantile of disease intensity losing between 5% and 20%; however, the coefficients are not statistically significant at usual levels of confidence. While time- and crop-invariant unobservables are controlled for, endogeneity and measurement error remain a concern. Since our measure of disease prevalence/intensity is model-based, the associated prediction error is included as a covariate, thereby hopefully reducing measurement error associated with the key RHS variable. However, since our intensity measure is constant at the village level, and because individual households may display heterogeneous levels of infection within a same village, measurement error in all likelihood remains. As such, we then proceed to instrument disease intensity with the density of the freshwater snails that serve as intermediate hosts of schistosomes. The third entry in Figure 2 shows how instrumenting intensity results in the point estimate of mean yield loss rising to 8.48% (40.99% in the top 5% intensity quantile). If measurement error is addressed by using heteroscedasticity as an internal instrument through the two-step GMM procedure of Lewbel (2012), the estimates remain very similar and are therefore omitted. In the Appendix we show how the results are robust to potential leptokurtic disturbances: the marginal effect of the disease at the median yield using the panel quantile procedure proposed by Koenker (2004) results in similar estimates.

Once we relax the functional form assumption on the production technology and adopt DML techniques, point estimates of the marginal effect of schistosomiasis intensity fall slightly, but precision improves. Indeed, as is apparent in Figure 2, all DML 95% confidence intervals are entirely contained within the confidence interval of the corresponding quasi-linear model, once household fixed effects are accounted for. For our 2009-2011 sample, our preferred non-instrumented model is given by gradient boosting machines, and results in a mean yield loss of 6.22% (30.05% in the top 5% intensity quantile). All three adaptive methods (random forests, gradient boosting machines and deep neural networks) perform similarly, and differences in MSE between the three are marginal. The IV DML estimates, which are computationally expensive in that they require at least 2-fold cross validation, yield slightly larger marginal effects, which are again estimated quite precisely. The instrumented DML estimate implies a mean loss of yield of 6.63% (32.04% in the top 5% intensity quantile).

An interesting characteristic of our dataset is the homogeneity of households in terms of yield: in Figure 3 we plot the density of yields in levels as well as the corresponding density after the within-village transformation. The densities are almost visually identical, and a Kolmogorov-Smirnov test cannot reject the null of equal distributions. A similar finding is presented by Udry (1996), using Burkina Faso ICRISAT survey data from 1981-1983. Furthermore, the yield density remains equally unchanged after partialling out household unobservables: plot-level heterogeneity, therefore, seems to be driven by plot characteristics. We then aggregate our data up to the village level: given the aforementioned homogeneity of household yields, village aggregation leads to what is essentially a representative household model.

Figure 2 Estimates of the loss in yield due to schistosomiasis for 2009 and 2011 (95% confidence interval). Each label reports average and top 5% percentile losses. All estimations include the full set of controls and time fixed effects, and plot-level estimations also include crop fixed effects. Errors are clustered at village level (commune level for commune fe), and cluster-bootstrapped for 2SLS.

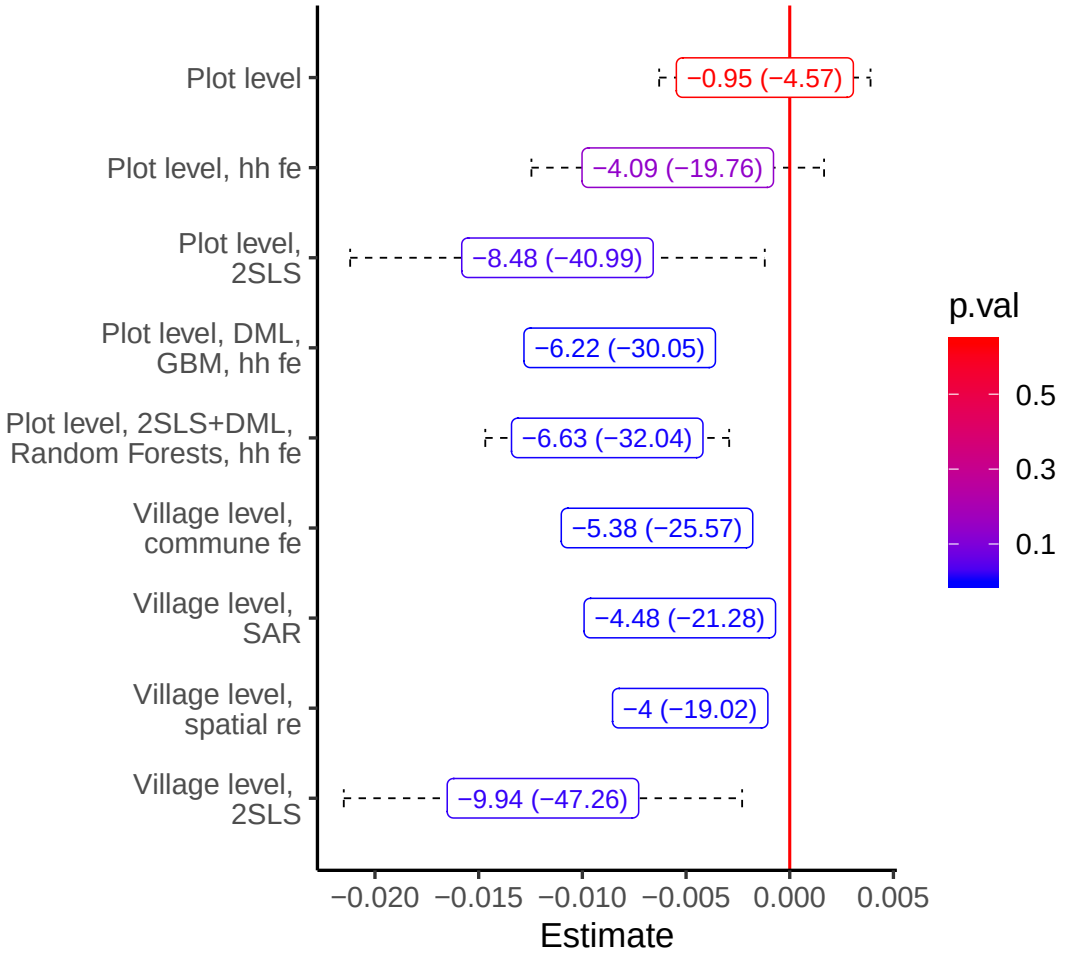
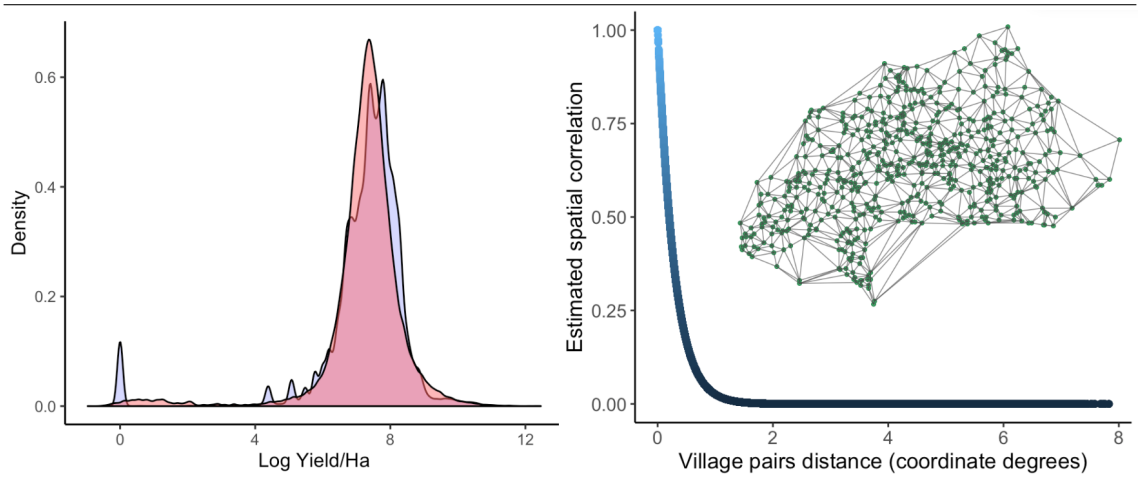


Figure 3 (Left) Homogeneity of households: original density of yield vs. rescaled density with village fixed effects partialled out. (Right) Distances between villages as spatial weights (inset), and the estimated spatial correlation.



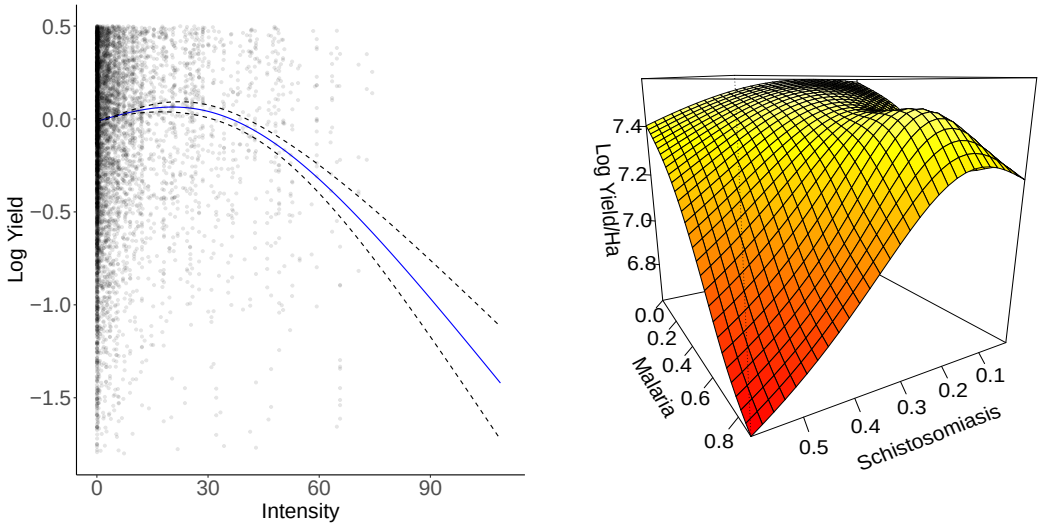
As expected, village-level estimates are more precise: within-cluster correlation is dampened and measurement error is reduced. Estimation of the linear model with commune fixed effects yields a loss in yield of 5.38% (25.57% in the top 5% intensity quantile), while the IV results yield a loss of 9.94% (47.26% in the top 5% intensity quantile). Aggregating at the village level also allows us to study the impact of spatial correlation, with the key issue being whether controlling for higher (commune) level unobservables is sufficient to control for spatial correlation. We begin by estimating Moran's I for the models with only time fixed effects as opposed to time and commune fixed effects: spatial correlation falls from 0.13 - low, but statistically significant at all conventional levels of confidence ($p = 2e^{-5}$) to a statistically insignificant -0.07 ($p = 0.970$). Commune level unobservables therefore seem to lie at the root of any spatial correlation present in the data. Fitting a spatial lag model with a weight matrix obtained from distances between villages yields a similar point estimate of 4.48%.[4] The map in the right-hand panel of Figure 3 provides a graphical illustration of the weight matrix. We then check for the presence of spatial random effects. Point estimates are slightly lower, with an average yield loss of 4%, but the null that this is equal to previous estimates cannot be rejected. We then construct a measure of spatial correlation from the parameters obtained from the spatial estimations. The right-hand panel of Figure 3 shows that spatial correlation decays rapidly: it vanishes almost completely at a distance of 1 degree (110 km at the equator), and by 0.33 degrees (ca. 37 km) it already falls below 0.1. This confirms the previous result that commune-level fixed effects already remove most spatial dependence in the data.

5.2 Non-linearity and interactions

The left-hand panel of Figure 4 presents semi-parametric results for our 2009-2011 sample. Results for other years, which are similar, are again relegated to the Appendix. There is a small *positive* effect of schistosomiasis on agricultural yield for low levels of intensity, with the significantly negative marginal effect beginning at an infection intensity of 30 eggs/person. This small positive marginal effect for low intensity stems from the interaction between schistosomiasis and malaria:

[4]The choice of queen or rook-type adjacency weights does not significantly affect the estimates.

Figure 4 (Left-hand panel) Nonlinear effect of schistosomiasis on log yield for households surveyed in 2009 and 2011, estimated via adaptive splines in a quasi-log-linear model. Each dot in the figure represents a household, upper portion of the data omitted for clarity of presentation. (Right-hand panel) Interaction of schistosomiasis and malaria prevalence measures.



given that malaria is only available as a prevalence measure, we fit a multidimensional spline for the interaction between malaria and schistosomiasis *prevalence*. The relationship is highly nonlinear, as displayed in the right hand panel of Figure 4: high prevalence levels for both diseases lead to a high negative impact on crop yields, but low to intermediate levels of schistosomiasis mitigate the negative effects of high malaria prevalence. The immunological mechanisms behind this phenomenon have been discussed by Hartgers and Yazdanbakhsh (2006) and Degarege et al. (2016), who present evidence on how low- to mid-intensity schistosomiasis can protect from acute malaria infections.

6 Schistosomiasis and water resources development

We begin by concentrating our attention on Burkina Faso's four main dams: the Bagré Dam in Boulgou province, which also involves Zoundwéogo, Kouritenga and Ganzourgou provinces, the Kompienga Dam in Kompienga province, the Ziga Dam in Oubritenga province and the Léry Dam in Nayala province (Fig. 1b). We are interested in whether the presence of a large dam affects the magnitude of the negative impact of schistosomiasis on agricultural yields: we therefore augment our initial model with an indicator variable $\mathbb{1}_{dam}$, representing whether a household is located in a province among those affected by the four large dams, as well as with the interaction between schistosomiasis intensity and the indicator. The coefficient associated with $I_{jt} \times \mathbb{1}_{dam}$ represents the difference in the effect of schistosomiasis on log yield attributable to the presence of a dam. This specification allows us to disentangle the direct impact of dams, which should increase yields, from the deleterious indirect effect that they may produce by facilitating the diffusion of schistosomiasis. The upper panel of Figure 5 reports the results: irrespective of estimation method, the interaction term is always statistically significant at conventional levels of confidence, and the presence of a large dam increases the average loss of agricultural yield due to schistosomiasis by a minimum of 8.2% to a maximum of 18.7%. The presence of a dam induces spatial effects that do not vanish,

contrary to our earlier results: this is confirmed by how Moran's I statistic for the village-level model, even after controlling for region-specific effects, is still significantly positive (0.10, highly significant). Indeed, it is only when a SAR framework is adopted that the separate effects of intensity and dams are estimated with sufficient precision to make them statistically significant: the presence of a dam increases yield by 23.8%, and exacerbates the negative effect of schistosomiasis by 9.9%. Plot level estimation yields mixed results: the interaction term remains negative and precisely estimated, while the intensity variable becomes statistically indistinguishable from zero whether one controls for regional fixed effect or applies the 2SLS procedure. The upshot is that our plot-level estimates suggest little effect of schistosomiasis outside of provinces with large dams, though our SAR results remain our preferred specification from the policy standpoint.

We refine the previous results by accounting for each village's distance (in km) from the nearest dam or water reservoir. We start by fitting an adaptive spline to the interaction of each village's distance from the dams and schistosomiasis intensity. The full network of dams and reservoirs is illustrated in panel (b) of Figure 1. Results are displayed in the middle panels of Figure 5: the left-hand panel shows how the deleterious marginal effect of schistosomiasis intensity on log yield is mitigated as one moves further away from a dam or a reservoir. The right-hand panel, based on the same estimation procedure, highlights how areas with high schistosomiasis intensity are concentrated within 20 km of a dam or reservoir. Households located in these areas suffer from large negative feedback effects between schistosomiasis and water resources development; to make matters worse, the effect of an increase in distance on the marginal effect of schistosomiasis intensity is greater for villages which display lower disease intensity. It is therefore of substantial relevance, particularly from a policy perspective, to investigate whether it is not only large-scale dams that are the culprits when it comes to the transmission of the disease, and whether smaller scale project built for livestock, irrigation and flood control can potentially generate as much of an adverse effect as large dams when it comes to the diffusion of the disease. To paraphrase the old adage: the road to Hell may be paved with good irrigation intentions. We therefore include the triple interaction $I_{jt} \times \mathbb{1}_{dam} \times dist_j$, as well as all the double interactions, at a village level by both region fixed effects and SAR: the results are presented in the lower panel of Figure 5 and are very similar. For the villages near any of the large dams, an increase of one kilometer in distance from the dam generates an average *reduction* of around 1.5% of the schistosomiasis burden on agricultural yield, starting from the average 18-20% loss suffered by the closest villages: being further away from large dams is therefore beneficial. The coefficient associated with the interaction $I_{jt} \times dist_j$ is always extremely small and insignificant at any reasonable level of confidence: large dams seem indeed to be the main culprits of the feedback effects. The consequences from the standpoint of poverty and inequality can be substantial: populations that gain the most from such large irrigation projects often do not correspond to those most exposed to their deleterious consequences in terms of health and productivity.

7 Schistosomiasis and poverty

All of the preceding estimates are compatible with a model in which schistosomiasis acts as a productivity shock without directly affecting the allocation of inputs. It is worth noting that all estimates include a measure of malaria prevalence which, interestingly, does not significantly affect yield at conventional levels of confidence. Schistosomiasis's impact, therefore, appears to be long-term and to affect total factor productivity. Furthermore, we will now show that plots where cotton is grown are both larger and seem to be immune to any deleterious effects of schistosomiasis on yield. There are potentially two reasons for this. First, cotton is relatively drought-resistant and

requires less by way of irrigation networks. Second, households farming cash crops such as cotton are on average much richer: they may therefore have readier access to clean running water and enjoy better sanitary conditions. In contrast, households that rely on subsistence agriculture suffer the most from the schistosomiasis productivity shock. Moreover, quantile regression estimates reveal that the marginal effect of schistosomiasis is indeed worse for less productive households, i.e. those in the lower quantiles in terms of log yield.[5] Poverty thus reinforces the negative economic impact of schistosomiasis. Conversely, development interventions that increase agricultural productivity and allow peasants to diversify into cash crops will both improve living standards and reduce the burden of the disease.

7.1 Schistosomiasis, crop choices and burden heterogeneity

We now investigate whether schistosomiasis affects households differently based on whether they cultivate cash crops (cotton) or food crops (all the rest, mostly of subsistence). The question is of substantial policy relevance: if schistosomiasis exerts a lower burden for cash crops, then policies aimed at poverty reduction and diversification of the agricultural sector will carry along a beneficial effect on the reduction of the economic impact of schistosomiasis. the lower panel of Fig.7 shows how plots that cultivate cotton as a cash crop are substantially larger than the others. Larger farms in Burkina Faso involve richer households, especially when agriculture is not anymore of the subsistence variety. Cotton, furthermore, is drought-resistant and mostly rainfall-fed, therefore less in need of plots being around large water reservoirs or networks. We expect the effect to be reduced for these crops: we also expect the productivity shock to be less harmful for farms and households where constraints are less binding and agricultural yield is not as directly crucial for survival. The upper panel of Figure 6 estimate the main plot-level model with the same techniques as the main estimation. Errors are again clustered at village level and cluster-boostrapped for instrumented regressions. Instrumentation of the model increases the estimates by a large amount, because subsetting to cotton crops reduces greatly the reservoir area of the snail hosts, as well as their estimated density: IV estimates are then large and extremely noisy. In any case, there does not seem to be a significant adverse effect of schistosomiasis on households farming cotton as cash crop, and the reason is likely to be twofold. The first reason is in cotton being a drought-resistant crop and thus requiring less irrigation networks: in the paper we show how this directly reduces the effective burden of the disease. The second reason lies in the fact that farms and households farming cash crops own substantially larger plots and are on average much richer; this can have multiple implications, from increased access to clean running water to better sanitary conditions. As the lower panel of Figure 6 shows, food crops, and particularly smaller plots mostly dedicated to farming of the subsistence variety, seem to suffer most of the disease burden, and show how schistosomiasis is indeed a disease of poverty. Economic development aimed at raising individuals and households from having to farm for survival to a standard of agriculture with better returns and conditions will generate both an improvement in living standards and a reduction of the disease burden.

Since yield may be subject to leptokurtic disturbances and the marginal impact of schistosomiasis may differ across yield quantiles, we also estimate conditional quantiles, using the panel quantile procedure proposed by Koenker (2004). We first obtain the marginal effect of schistosomiasis on the median yield: we obtain a median effect of -4.48%, consistent result with the main linear panel estimations. This is unsurprising, given that the log yield mean is close to the log yield median, but adds robustness to the result, as it's not anymore relying on Gaussian errors. The estimate and median burden of the disease is shown as the 50th quantile in the lower panel of

[5] See Koenker (2004) for computational details and the Appendix for the estimates.

Figure 5 (Upper panel) Differential effect of schistosomiasis on yields caused by the presence of a large dam. (Middle panels) Estimated joint impact of schistosomiasis intensity and distance from all dams and water reservoirs. (Lower panel) Joint effect of schistosomiasis and distance from a large dam. All estimates control for time fixed effects, and plot-level estimates control for crop fixed effects. Errors are clustered at the regional level for region fixed effects estimates, and at the village level otherwise. 2SLS standard errors are cluster-bootstrapped.

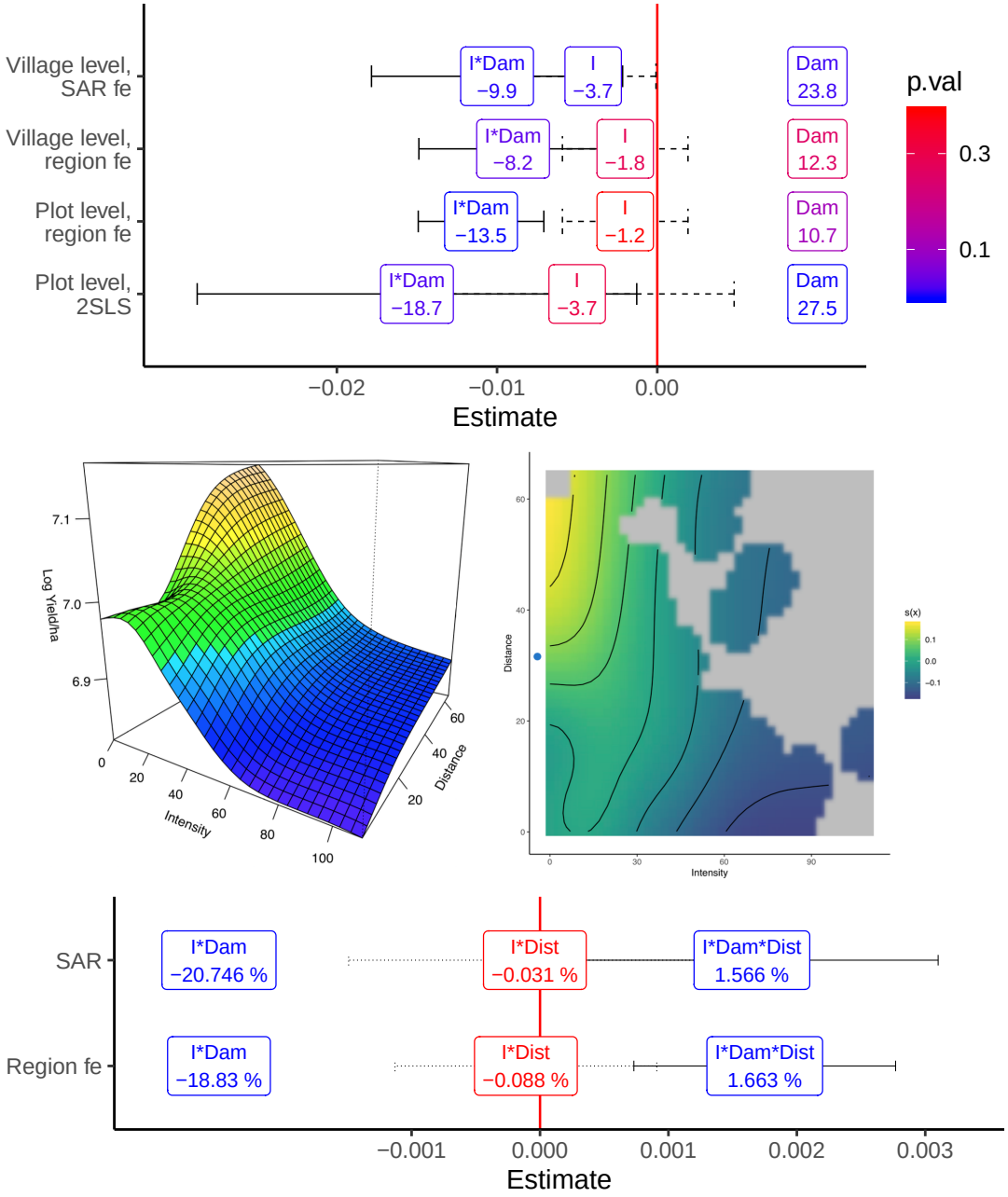


Figure 6 (Upper panel) Schistosomiasis burden for cash crops-only plots: no effect
(Middle panel) Burden for food crops, mostly of subsistence: large effect. In labels: percentage yield loss for villages at the average schistosomiasis intensity
(Lower panel) Burden of schistosomiasis on different conditional log yield quantiles (5%, 10%, 25%, 50%,75%, 90%).

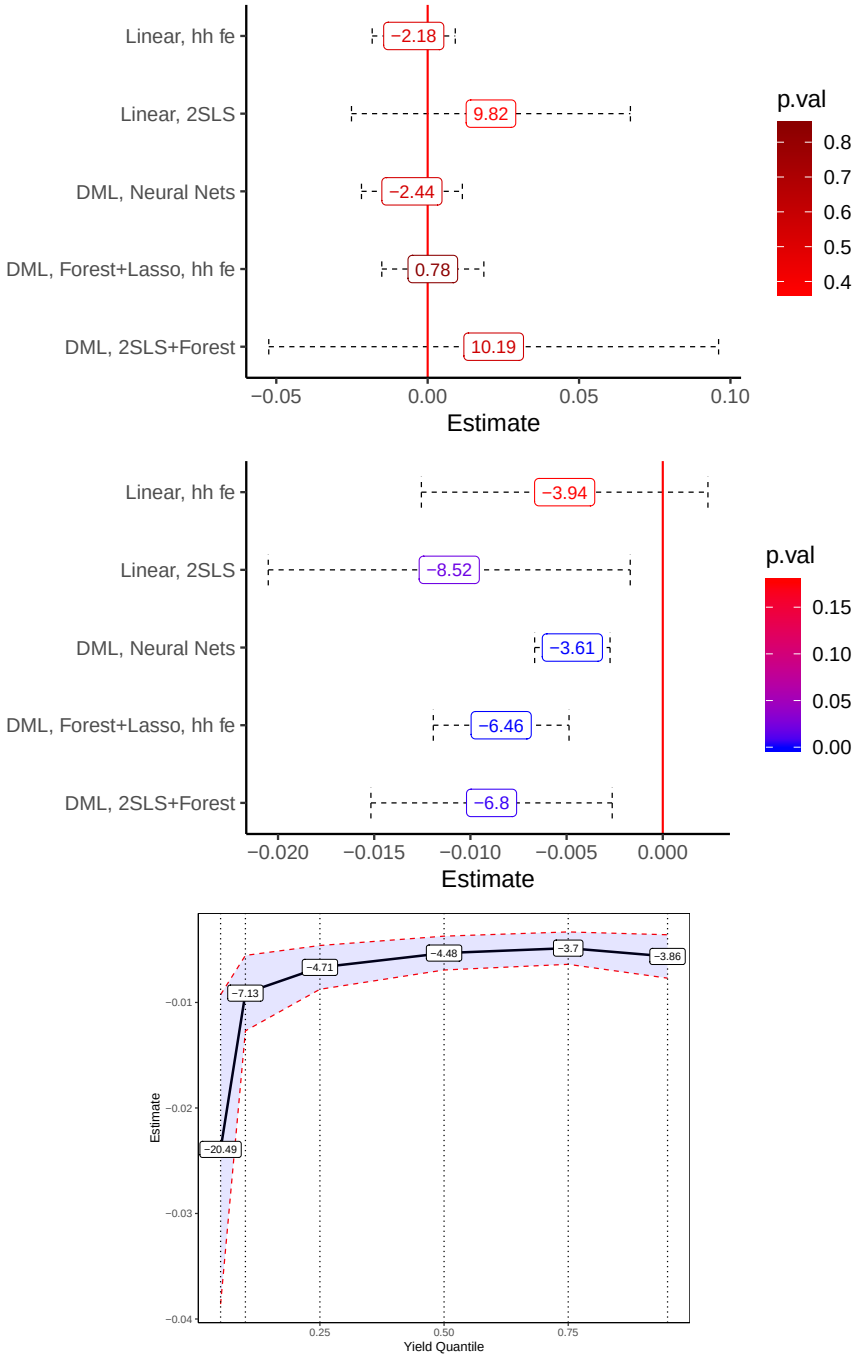


Figure 6, while the estimates of the other determinants of agricultural yield are reported in 8 of the Appendix . There is considerable heterogeneity in the disease burden across households who farm plots with different yields: the lower panel of Figure 6 shows how the disease burden for the households at the lowest 5% yield quantile is drastically larger (-20.49%), and significantly lower than the burden suffered by households at the upper quantiles. Households at the lowest 10% yield quantile also suffer a higher burden (-7.13%) than the rest, although the effect is not as strikingly large as for the ones at the lowest 5%. This result shows that schistosomiasis is indeed a disease of poverty.

8 Conclusions

In this paper we have studied the impact of schistosomiasis on agricultural yields in Burkina Faso by characterizing it as a productivity shock. By merging rich plot and household datasets with high-resolution disease maps, we have shown how schistosomiasis has a large, negative and nonlinear effect on agricultural productivity. Losses attributable to the disease range from a mean value of 6.6% to 32% for households and villages located in areas in the top 5% of schistosomiasis intensity: furthermore, such losses are concentrated on non-cash crops, mostly used for subsistence. From the policy perspective, perhaps our most interesting result is that while dams and reservoirs, *ceteris paribus*, increase agricultural yields, they can also induce substantial negative feedback effect by spreading the disease and furthering its deleterious impact on agricultural productivity. We show that this feedback is generated entirely by large-scale dams. Our work highlights how the study of the interactions between disease diffusion and economic development can benefit from the use of high-resolution data, which allows one to control for the numerous confounding factors that aggregated data necessarily misses. While our focus has been on Burkina Faso, in part because it is likely to be a worst-case scenario, our approach can be applied to any country in which schistosomiasis in endemic and, indeed, to all other water-borne diseases.

References

Acemoglu, D. and S. Johnson (2007). Disease and development: The effect of life expectancy on economic growth. *Journal of Political Economy 115*(6), 925–985.

Acemoglu, D., S. Johnson, and J. Robinson (2003). Disease and development in historical perspective. *Journal of the European Economic Association 1*(2-3), 397–405.

Adenowo, A. F., B. E. Oyinloye, B. I. Ogunyinka, and A. P. Kappo (2015). Impact of human schistosomiasis in sub-saharan africa. *The Brazilian Journal of Infectious Diseases 19*(2), 196 – 205.

Anderson, R. and G. Medley (1985). Community control of helminth infections of man by mass and selective chemotherapy. *Parasitology 90*(4), 629–660.

Andrade, G., D. J. Bertsch, A. Gazzinelli, and C. H. King (2017). Decline in infection-related morbidities following drug-mediated reductions in the intensity of schistosoma infection: A systematic review and meta-analysis. *PLoS Neglected Tropical Diseases 11*(2), e0005372.

Audibert, M. (1986). Agricultural non-wage production and health status: A case study in a tropical environment. *Journal of Development Economics 24*(2), 275–291.

Audibert, M. (2010). Endemic diseases and agricultural productivity: Challenges and policy. *Journal of African Economies 19*(3), 110–165.

Audibert, M. and J.-F. Etard (1998, 07). Impact of Schistosomiasis on Rice Output and Farm Inputs in Mali. *Journal of African Economies 7*(2), 185–207.

Baldwin, R. E. and B. A. Weisbrod (1974). Disease and labor productivity. *Economic Development and Cultural Change 22*(3), 414–435.

Belloni, A. and V. Chernozhukov (2013, 05). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies 81*(2 (287)), 608–650.

Bleakley, H. and F. Lange (2009). Chronic disease burden and the interaction of education, fertility, and growth. *The Review of Economics and Statistics 91*(1), 52–65.

Boelee, E., P. Cecchi, and A. Koné (2009). *Health impacts of small reservoirs in Burkina Faso*. International Water Management Institute (IWMI).

Brooker, S., S. Whawell, N. B. Kabatereine, A. Fenwick, and R. M. Anderson (2004). Evaluating the epidemiological impact of national control programmes for helminths. *Trends in parasitology 20*(11), 537–545.

Cecchi, P., A. Meunier-Nikiema, N. Moiroux, B. Sanou, and F. Bougaire (2007). Why an atlas of lakes and reservoirs in Burkina Faso? Technical Report iii, Small Reservoirs Project.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Colley, D. G., A. L. Bustinduy, W. E. Secor, and C. H. King (2014). Human schistosomiasis. *The Lancet 383*(9936), 2253–2264.

Degarege, A., D. Degarege, E. Veledar, B. Erko, M. Nacher, C. M. Beck-Sague, and P. Madhivanan (2016, 12). Plasmodium falciparum infection status among children with schistosoma in sub-saharan africa: A systematic review and meta-analysis. *PLOS Neglected Tropical Diseases 10*(12), 1–18.

Diakité, N. R., M. S. Winkler, J. T. Coulibaly, N. Guindo-Coulibaly, J. Utzinger, and E. K. N'Goran (2017). Dynamics of freshwater snails and *Schistosoma* infection prevalence in schoolchildren during the construction and operation of a multipurpose dam in central Côte d'Ivoire. *Infectious Diseases of Poverty 6*(1), 93.

Didan, K. (2015). Mod13a2 modis/terra vegetation indices 16-day l3 global 1km sin grid v006 [data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-03-06 from https://doi.org/10.5067/MODIS/MOD13A2.006.

Ezeamama, A. E., A. L. Bustinduy, A. K. Nkwata, L. Martinez, N. Pabalan, M. J. Boivin, and C. H. King (2018). Cognitive deficits and educational loss in children with schistosome infectionâĂŤa systematic review and meta-analysis. *PLoS Neglected Tropical Diseases 12*(1), e0005524.

Foster, R. et al. (1967). Schistosomiasis on an irrigated estate in east africa. iii. effects of asymptomatic infection on health and industrial efficiency. *Journal of Tropical Medicine and Hygiene 70*(8), 185–195.

Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, et al. (2015). The climate hazards infrared precipitation with stationsâĂŤa new environmental record for monitoring extremes. *Scientific data 2*(1), 1–21.

Hartgers, F. C. and M. Yazdanbakhsh (2006). Co-infection of helminths and malaria: modulation of the immune responses to malaria. *Parasite Immunology 28*(10), 497–506.

Hastie, T., R. Tibshirani, J. Friedman, et al. (2001). The elements of statistical learning.

James, S. L., D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim, et al. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet 392*(10159), 1789–1858.

King, C. H. (2010). Parasites and poverty: The case of schistosomiasis. *Acta Tropica 113*(2), 95 – 104.

King, C. H. and M. Dangerfield-Cha (2008). The unacknowledged impact of chronic schistosomiasis. *Chronic illness 4*(1), 65–79.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74 – 89. Special Issue on Semiparametric and Nonparametric Mixed Models.

Lai, Y.-S., P. Biedermann, U. F. Ekpo, A. Garba, E. Mathieu, N. Midzi, P. Mwinzi, E. K. N'Goran, G. Raso, R. K. AssarÃČÂľ, M. Sacko, N. Schur, I. Talla, L.-A. T. TchuentÃČÂľ, S. TourÃČÂľ, M. S. Winkler, J. Utzinger, and P. Vounatsou (2015). Spatial distribution of schistosomiasis and treatment needs in sub-saharan africa: a systematic review and geostatistical analysis. *The Lancet Infectious Diseases 15*(8), 927 – 940.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1), 67–80.

Mbabazi, P. S., O. Andan, D. W. Fitzgerald, L. Chitsulo, D. Engels, and J. A. Downs (2011). Examining the relationship between urogenital schistosomiasis and hiv infection. *PLoS Neglected Tropical Diseases 5*(12).

Miguel, E. and M. Kremer (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Ouedraogo, H., F. Drabo, D. Zongo, M. Bagayan, I. Bamba, T. Pima, F. Yago-Wienne, E. Toubali, and Y. Zhang (2016). Schistosomiasis in school-age children in burkina faso after a decade of preventive chemotherapy. *Bulletin of the World Health Organization 94*(1), 37.

Perez-Saez, J., T. Mande, N. Ceperley, E. Bertuzzo, L. Mari, M. Gatto, and A. Rinaldo (2016). Hydrology and density feedbacks control the ecology of intermediate hosts of schistosomiasis across habitats in seasonal climates. *Proceedings of the National Academy of Sciences 113*(23), 6427–6432.

Perez-Saez, J., T. Mande, J. Larsen, N. Ceperley, and A. Rinaldo (2017). Classification and prediction of river network ephemerality and its relevance for waterborne disease epidemiology. *Advances in water resources 110*, 263–278.

Perez-Saez, J., T. Mande, and A. Rinaldo (2019). Space and time predictions of schistosomiasis snail host population dynamics across hydrologic regimes in burkina faso. *Geospatial health 14*(2).

Perez-Saez, J., L. Mari, E. Bertuzzo, R. Casagrandi, S. H. Sokolow, G. A. De Leo, T. Mande, N. Ceperley, J.-M. Froehlich, M. Sou, H. Karambiri, H. Yacouba, A. Maiga, M. Gatto, and A. Rinaldo (2015, 10). A theoretical analysis of the geography of schistosomiasis in burkina faso highlights the roles of human mobility and water resources development in disease transmission. *PLOS Neglected Tropical Diseases 9*(10), 1–27.

Poda, J., A. Traoré, and B. K. Sondo (2004). L'endémie bilharzienne au burkina faso. *Societé de Pathologie Exotique 97*(1), 47–52.

Poda, J.-N., S. P. Wango, H. Sorgho, and D. Dianou (2004). Évolution récente des schistosomoses dans le complexe hydroagricole du Sourou au Burkina Faso. *Société de Pathologie Exotique 1*(97), 15–18.

Richter, J. (2003). The impact of chemotherapy on morbidity due to schistosomiasis. *Acta Tropica 86*(2-3), 161–183.

Steinmann, P., J. Keiser, R. Bos, M. Tanner, and J. Utzinger (2006). Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases 6*(7), 411 – 425.

Udry, C. (1996). Gender, agricultural production, and the theory of the household. *Journal of Political Economy 104*(5), 1010–1046.

Wan, Z., S. Hook, and G. Hulley (2015). Mod11a1 modis/terra land surface temperature/emissivity daily l3 global 1km sin grid v006 [data set]. NASA EOSDIS Land Processes DAAC. Accessed 2020-03-06 from https://doi.org/10.5067/MODIS/MOD11A1.006.

Weisbrod, B. A., R. L. Andreano, R. E. Baldwin, E. H. Epstein, A. C. Kelley, and T. W. Helminiak (1974). Disease and economic development: The impact of parasitic diseases in st. lucia. *International Journal of Social Economics 1*(1), 111–117.

World Bank Results (2017). Agriculture as a powerful instrument for poverty reduction.

Appendices

A Agricultural and Household data

The agricultural and household surveys we use has been obtained from the Direction GŃérale des Previsions et de Statistiques Agricoles at the Institut National de la Statistique et la Démographie in Ouagadougou (Burkina Faso), ranging 2003 to 2017. The agricultural dataset is at plot level while many of the variables in the household survey are at individual level. Because of the nature of the analysis, we have chosen to maintain all the information at plot level and aggregate individual variables in the household-level dataset at a household level, to then merge together via unique household identifiers into a large plot-level dataset. Overall the dataset contains 19,993 unique households located in 1,950 villages (see Fig.(1a) in the paper), cultivating 202,162 unique plots over the years. Each household on average cultivates around 10 plots throughout the years. Each plot can have up to two different crops, among 31 possible ones. The most common food crops are millet, corn, rice, sorghum and peanuts, and the only cash crop is cotton (traditional, bio, GMO). Figure 7 shows the counts of crop choices per yield in the 2009-2011 dataset. Most crops are food crops, and cash crops are almost only chosen as primary crop. Plots that cultivate cash crops are substantially larger than the rest: the bottom panel of Figure 7 shows the density of the log surface of plots used for cotton and the overall surface density. As shown in Figure 3 of the paper, the distribution of plot yield is independent of household fixed effects, implying heterogeneity is substantially at plot level. Figure 1 (a) in the paper shows how the villages cover substantially the area of Burkina Faso, and allow us to exploit a large variation in terms of both disease prevalence and geographical heterogeneity. Households are not identified geographically, and we are able to geolocalize the data only at a village level. Tables 1 and 2 show which households and villages are repeated across the 2010 cutoff, and we choose 2009 and 2011 for robustness, number of observed crops and model fit. The first four plots of Figure 9 present the result of spline fitting for different year combinations and their consistency. The results in the main paper are for the two years chosen because of the best residual fit (bottom panel). Figure 8 presents the maps of the villages for the five relevant years around the 2010 cutoff (2009 to 2013) and their aggregated average yield.

Table 1 Repeated households

	2011	2012	2013	2014	2015	2016	2017
2003	4	4	4	2	2	2	2
2005	5	6	6	3	3	3	3
2006	5	6	6	2	2	2	2
2007	5	6	6	3	3	3	3
2008	119	120	128	19	18	18	19
2009	3,040	2,994	3,185	10	10	10	11
2010	3,049	3,066	3,208	18	19	18	17

B Climatic and environmental covariates

Table 3 gives details on the data sources used in our analysis, and yearly means are shown in Figure 10.

Figure 7 (Upper two panels) Crop choices for plot cultivations (above: main crop, lower: second crop, 2009 and 2011). (Lower panel) Distribution of plot sizes for cotton crops vs. all crops

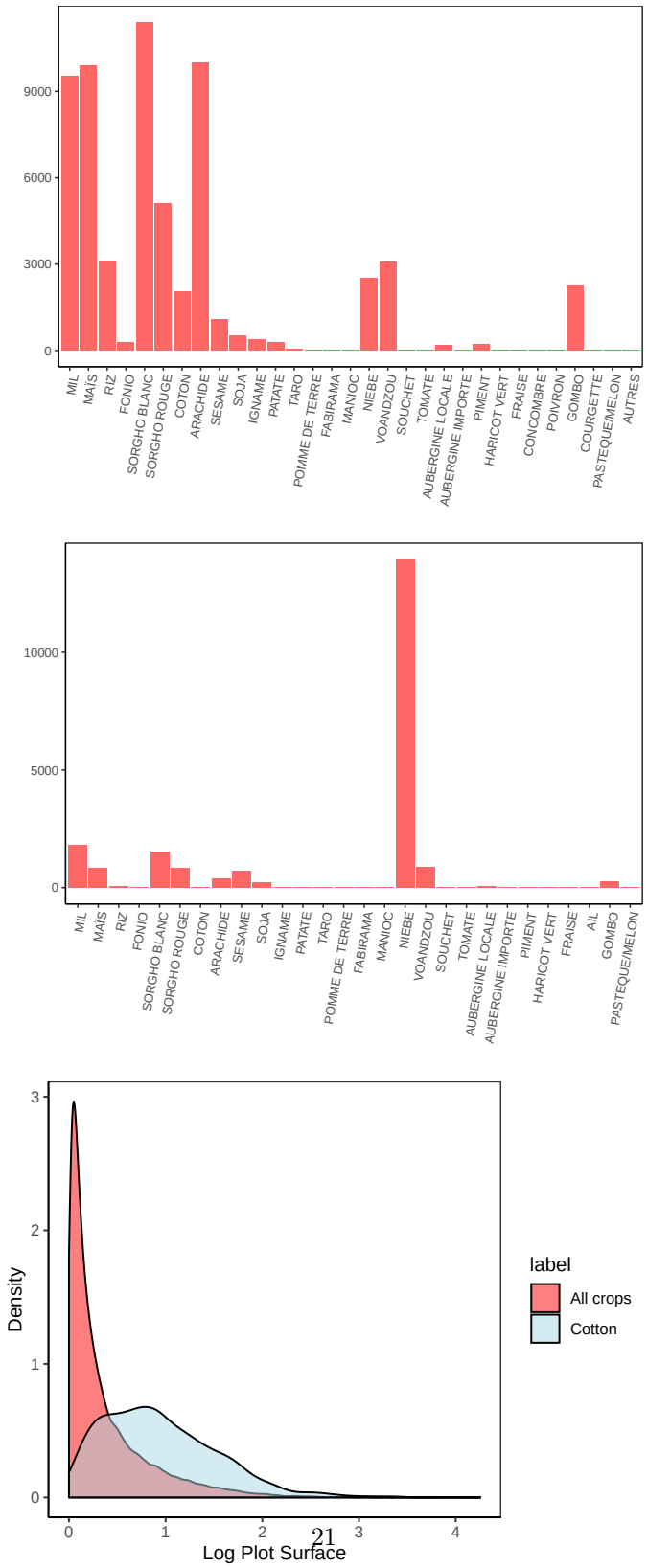


Figure 8 Geolocalized villages and relative average agricultural yield for the years 2009 to 2013 (from top left, in order)

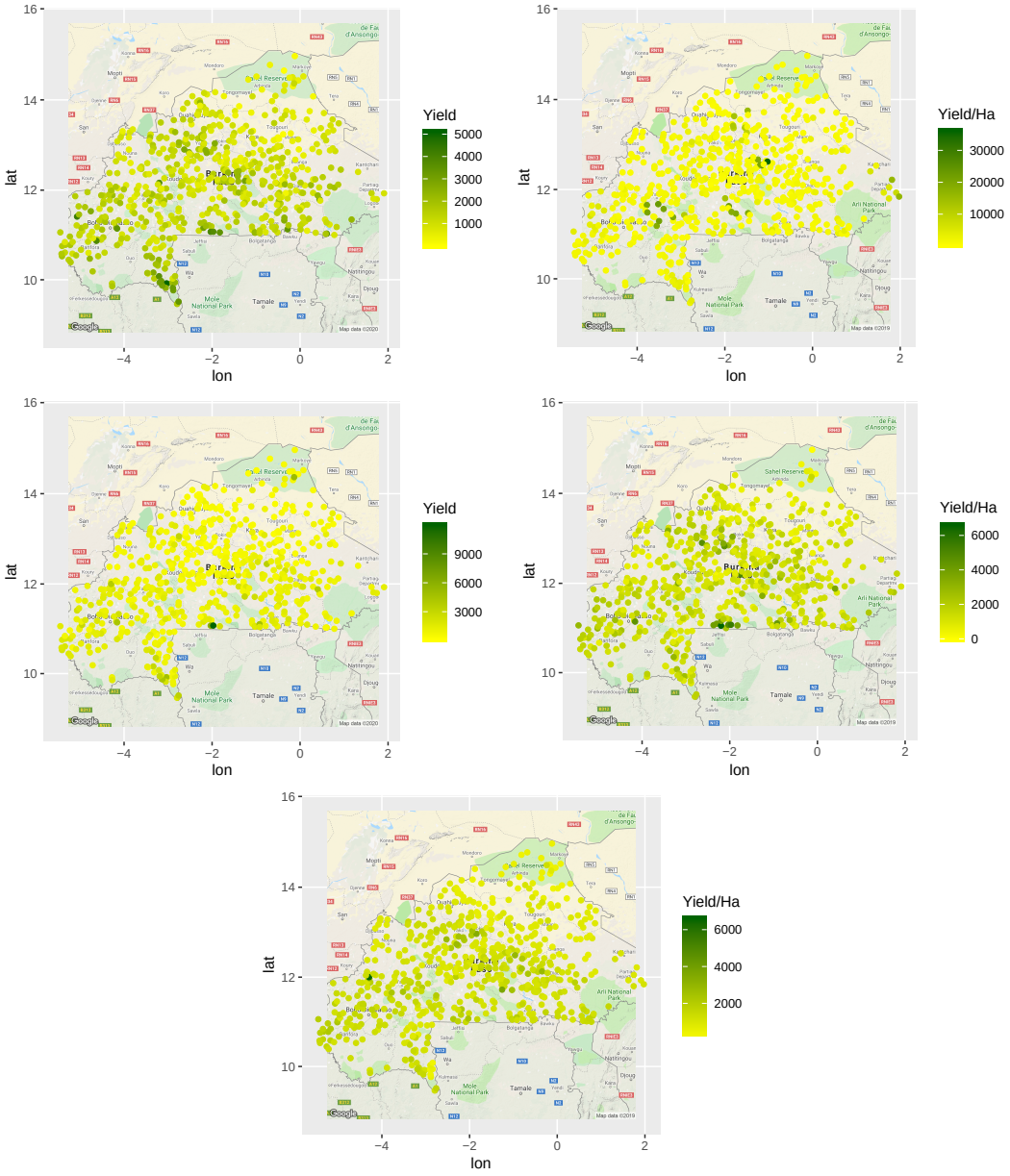


Figure 9 (top) Spline fit for 2009, 2012 and 2013.
(middle) Spline fit for 2010, 2013, and using all years.
(bottom) Residuals fit for the 2009-2011 years.

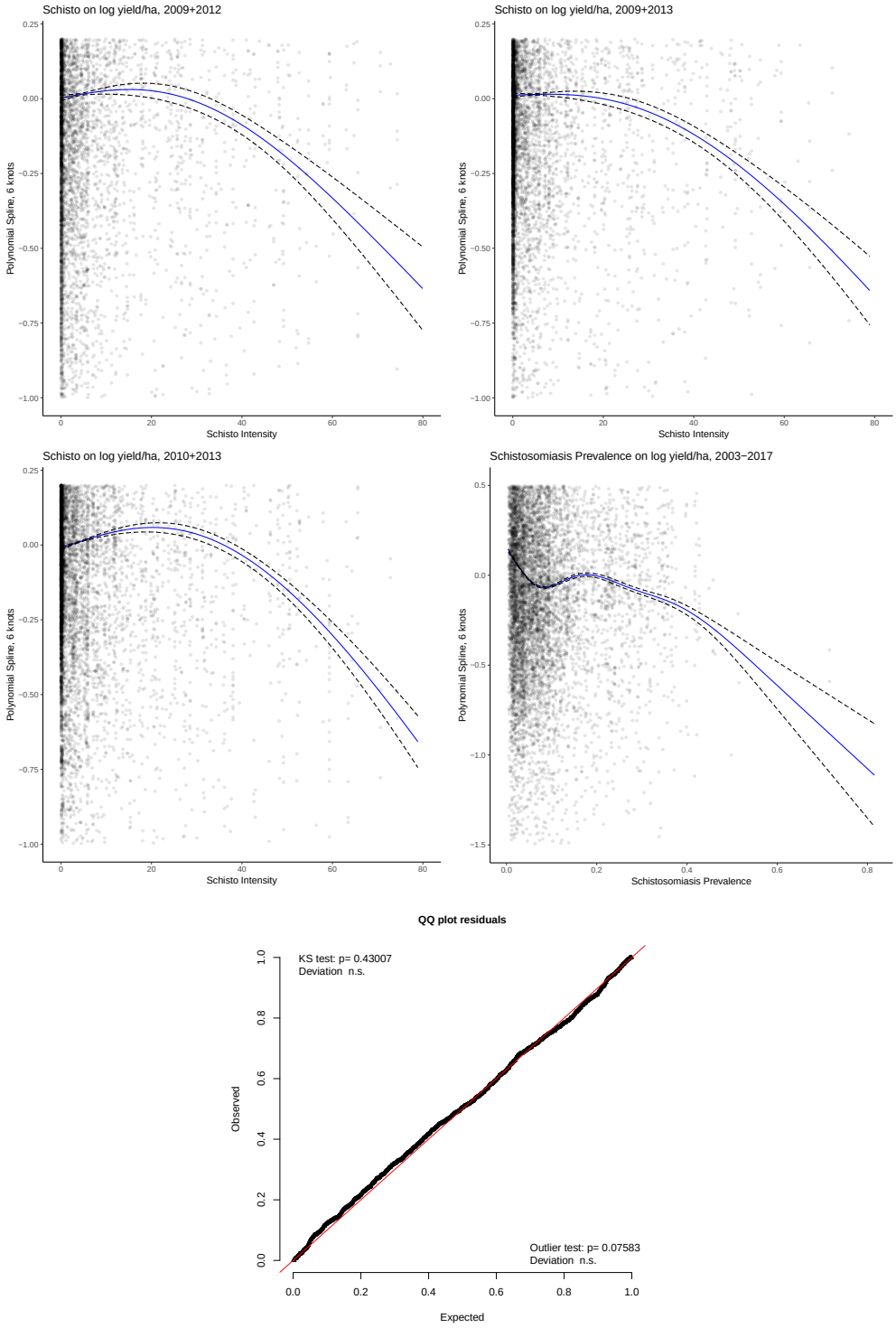


Figure 10 Climatic and environmental covariates used in the analysis. (a) Mean yearly precipitation, (b) mean dry-period precipitation (January to May), (c) Mean wet period precipitation (June to October), (d) Mean length of dry spells during the wet period. (e) Mean of maximum length of dry spells during the wet period, (f) Mean dry-period air temperature, (g) Mean day temperature, (h) Mean night temperature, (i) Mean Enhanced Vegetation Index (EVI). All figures show the yearly mean for the 2004-2018 period.

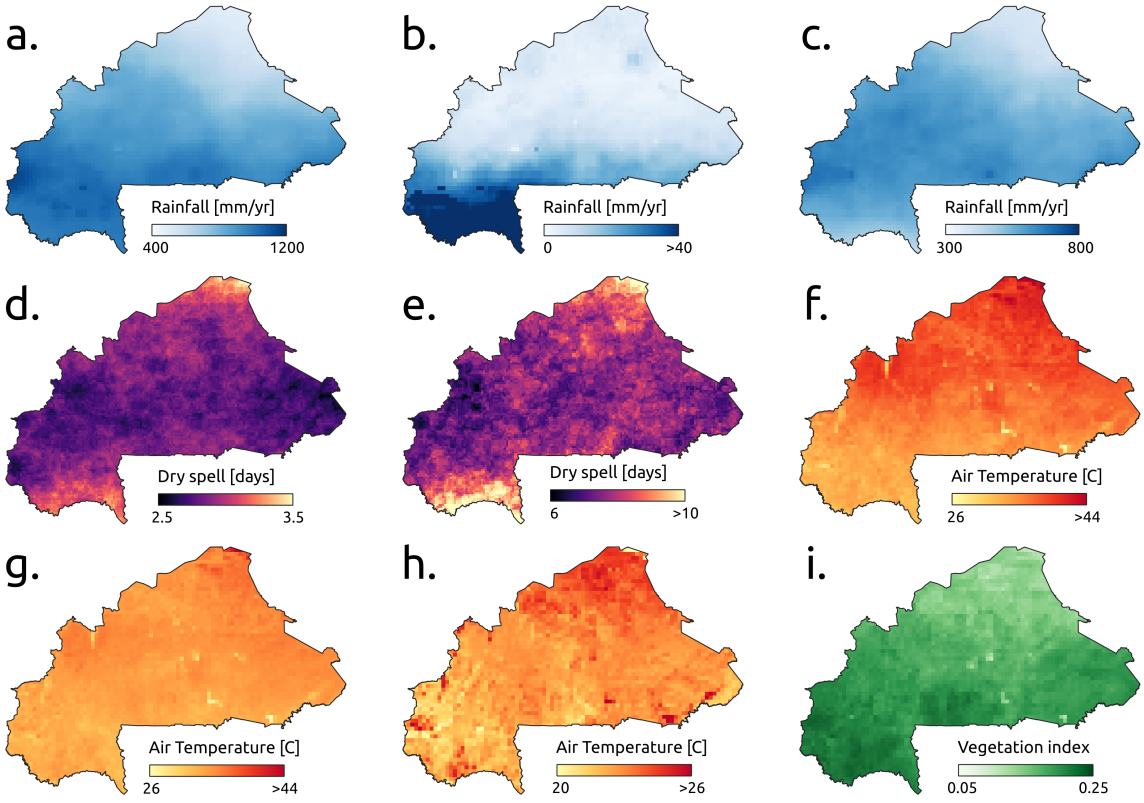


Table 2 Repeated villages

	2011	2012	2013	2014	2015	2016	2017
2003	18	18	19	16	16	16	16
2005	25	25	26	23	23	23	23
2006	25	25	26	21	21	21	21
2007	26	26	27	23	23	23	23
2008	651	648	697	235	233	235	236
2009	617	616	657	133	131	133	134
2010	588	586	618	125	124	125	126

Table 3 Details on remote sensing climatic and environmental covariates used in the analysis

Variable	Product	Source	Derived statistics
Daily rainfall	CHIRPS-V2	(Funk et al., 2015)	Yearly rainfall Dry-period rainfall Wet-period rainfall Mean length of dry spells Max length of dry spells
Monthly mean surface temperature	MODIS MOD11A1	Wan et al. (2015)	Mean yearly surface temperature
Monthly mean air temperature	-	Perez-Saez et al. (2019)	Mean yearly air temperature Mean yearly night temperature Mean dry-period day temperature
Vegetation indices	MODIS MOD13A2	Didan (2015)	Mean yearly Normalized Vegetation Index (NDVI) Mean yearly Enhanced Vegetation Index (EVI)

C Transforming schistosomiasis prevalence into infection intensity

The input to our analysis consists of modelled estimates of schistosomiasis prevalence Lai et al. (2015). Disease burden is however more directly related to infection intensity (worms per person), rather than prevalence. We therefore proceed to transforming modelled prevalence into infection intensity using common assumptions about the distribution of worm burden in the human population.

The data consists of prevalence and mean intensity of infection from parasitological surveys in Burkina Faso coming from the Schistosomiasis control program of the Ministry of Health, of which part of the data is published in (Ouedraogo et al., 2016). All samples correspond to school-aged children. We here make the assumption that the number of *Schistosoma* eggs per sample in the population follows a negative binomial distribution (Anderson and Medley (1985); Brooker et al. (2004)). We have that the prevalence is given by: $p = 1 - P(X = 0) = 1 - \left(1 + \frac{\mu}{k}\right)^{-k}$. Parasitological data across endemic countries suggest that the aggregation parameter k is not constant across transmission settings, but rather varies as a function of the mean intensity of infection μ in the population (mean eggs/person) (Brooker et al., 2004). Following other studies, we test for either a constant, linear ($k(\mu) = a + b\mu$) or a quadratic ($k(\mu) = a + b\mu + c\mu^2$) relation. Inference on $k(\mu)$ is drawn through maximum likelihood estimation. For a given parameters set θ, the likelihood of the parameters the parasitological data is given by a binomial distribution on the number of infected people $n_{infected,i}$ among the sampled population $n_{sampled,i}$ in each village i among m sampled villages:

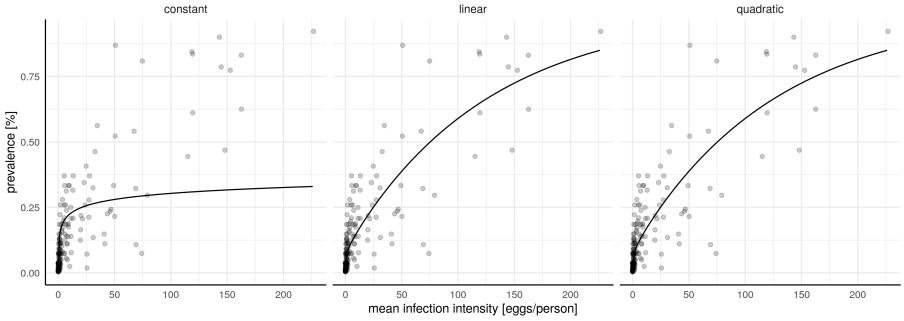
$$\mathcal{L}(\theta, \mathcal{D}) = P(\mathcal{D}|\theta) = \prod_{i}^{N} p_i^{n_{infected,i}} (1 - p_i)^{n_{sampled,i} - n_{infected,i}},$$

where p_i is the prevalence of schistosomiasis in village i. The negative log-likelihood is then minimized using standard optimization algorithms. We select for the best model using Akaike Information Criterion (AIC) AIC $= -2\log(\mathcal{L}) + 2m$, where m is the number of parameters in the model. We find the that there is a strong support for a non-constant functional form (Table 4). The best-fitting quadratic parameter has a value close to 0, so the linear form is retained.

Table 4 Model selection results for the functional form of the aggregation parameter for the inversion of schistosomiasis prevalence into infection intensity.

$k(\mu)$	$\log(\mathcal{L})$	AIC	parameters		
			a	b	c
constant	-6682.02	13366	0.047	-	-
linear	-6020.48	12043	0.016	0.0012	-
quadratic	-6020.48	12043	0.016	0.0012	1e-12

Figure 11 Functional forms of the variation of parasite aggregation with mean infection intensity in Burkina Faso. The best fitting relationship is a linear function of mean egg-intensity (Table 4).



D Schistosomiasis and optimal input allocation

We now present a simple model of optimal household input allocation with a distortion to labor input due to schistosomiasis. We posit the disease to be a shock to *effective* labor supply E, of the form

$$E = (\phi^{-\tilde{\theta}} L),$$

with $\theta > 0$, and that the total agricultural output of the household is generated by

$$Y = A(\phi^{-\tilde{\theta}} L)^{\alpha} F(X, K),$$

where A is total factor productivity, $\alpha < 1$ and $F(.)$ is the remainder of the production technology, dependent on k inputs$X$ chosen by the household (pesticides, livestock use) and inputs K which affect production but cannot be chosen (climate, land type). A household aims to maximize its instantaneous agricultural profits by choosing optimally$L$ and X, and is given by

$$\arg\max_{L,X} \Pi = p^y A(\phi^{-\tilde{\theta}}L)^{\alpha}F(X,K) - wL - \sum_{i=1}^{k} p_i X_i \tag{7}$$

where p_i are the input prices, w is the labor wage, and p^y is the price of output. The $k+1$ optimal allocations are then obtained by solving the system of $k+1$ first-order conditions:

$$L^*(A,\phi^{-\tilde{\theta}},K,w,p), X_i^*(A,\phi^{-\tilde{\theta}},K,w,p) \rightarrow \begin{cases} L^* = \left(\frac{\alpha p^y A F(X^*,K)}{w\phi^{\alpha\tilde{\theta}}}\right)^{\frac{1}{1-\alpha}} \\ X_i^* = F^{-1}\left(\frac{p_i\phi^{\alpha\tilde{\theta}}}{AL^*}\right) \end{cases}$$

where $p = (p^y, p_1, \ldots, p_k)$ is the vector of prices. Assuming an interior maximum, depending on the properties of $F(.)$, the log-linearized optimal production is therefore given by

$$y^* = \tilde{A} - \theta\phi + \alpha L^*(A,\phi^{-\tilde{\theta}},K,w,p) + \tilde{F}(X_i^*(A,\phi^{-\tilde{\theta}},K,w,p),K),$$

which recovers the quasilinear structure of the equation (2) of the paper. If F is assumed Cobb-Douglas, then the model becomes fully linear.

The total marginal effect of schistosomasis on log-linearized optimal production, therefore, is given by

$$\frac{dy^*}{d\phi} = -\theta + \alpha\frac{dL^*}{d\phi} + \sum_{i=1}^{k}\frac{\partial F(X^*,K)}{\partial X_i^*}\frac{dX_i^*}{d\phi},$$

where $\theta = \alpha\tilde{\theta}$, and the specification becomes identical to Eq. (1) (i.e. schistosomiasis is a pure productivity shock) if $\frac{dL^*}{d\phi}$ and all the $\frac{dX_i^*}{d\phi}$ are zero. Note that this model can straightforwardly be extended to a more general form $F(L,X,K)$, but we choose this specification to better illustrate the choice of estimating Eq. (1) and choosing both log-linear and adaptive forms for F.

The condition of zero derivatives is what we observe in the data: a first set of reduced-form evidence comes from studying the interactions of schistosomiasis on all input variables. We obtain a clear sign of zero signal: Table 6 reports the coefficients of the interactions, and omits both the rest of the controls as well as the interaction of schistosomiasis intensity on non-chosen inputs (climate, land type). Table 5 shows how for all of the choice input variables the marginal impact of schistosomiasis is not significant, hence validating our initial assumption of identifying schistosomiasis as a productivity shock in estimating Eq.(1). The results are robust to more flexible specifications (interactions, adaptive methods), as well as by instrumenting the intensity with the snail densities. All other controls used for each line of5 are omitted. Unlike schistosomiasis, malaria does not seem to have a significant effect in this specification as a productivity shock, which we find to be a believable finding: malaria, by the nature of its health effects and burdens, is more consistent with a temporary shock to direct labor supply. This identification would have to be tested with a completely different strategy, which we leave to future research.

Table 5 Input use and schistosomiasis intensity

y	*Intensity (Clustered s.e.)*
log plot surface	0.0003 (0.0002)
presence of hired labor	0.0003(0.0004)
presence of motorized labor	0.0008(0.0008)
presence of manual labor	0.0002(0.001)
presence of horse-driven labor	-0.001 (0.0007)
presence of labor from mutual aid	-0.002(0.001)
total agricultural livestock	0.0001(0.001)
total livestock	0.0001(0.0008)
cows	-0.0011(0.0021)
horses	-0.0024(0.0039)
pigs	-0.0014(0.0028)
goats	0.001(0.0021)
self-consumption	0.0018(0.0014)
cows sold	0.0002(0.0015)
cows bought	-0.0042(0.0028)
pigs sold	0.0004(0.002)
pigs bought	0.0008(7e-04)
goats sold	0.0002(5e-04)
goats bought	-0.0001(0.0012)
gifts	0.0001(0.001)
thefts	0.0011(0.001)
type of plot farming	-3e-04(0.0012)
npk (kg)	0.0023(0.0016)
urea (kg)	-0.0025(0.0015)
phosphates (kg)	0.0006(0.0006)
solid pesticides (kg)	0.0015 (0.001)
liquid pesticides (cl)	0.0014(0.0012)
herbicides (g)	0.0025(0.0036)
herbicides (cl)	-0.0014(0.0032)
fungicides (g)	-0.0017(0.0023)
fungicides (cl)	0.0008(0.0006)
rodenticides (g)	0.0036(0.0025)
rodenticides (cl)	0(0.0013)
number of household members	-0.011(0.0042)
number of families in the household	0.0006(0.0007)
number of young members working	0.0045(0.0026)
number of women working	0.0052 (0.0029)
rainfed agriculture	0.0001(0.0001)
market garden	0.0001(0.0011)
fishing	0.001(0.0008)
Observations	51,398
Time + household f.e., village clustering	*p<0.1; **p<0.05; ***p<0.01

Table 6 No signal on schistosomiasis interactions for choice inputs

	y: Log Yield
intensity×log surface	−0.001 (0.003)
intensity×presence of hired labor	−0.005 (0.003)
intensity×presence of motorized labor	0.008(0.008)
intensity:×presence of manual labor	−0.012(0.011)
intensity×presence of labor from mutual aid	0.002 (0.003)
intensity×total agricultural livestock	0.001(0.001)
intensity×total livestock	0.00000(0.00004)
intensity×cows	0.0003 (0.0003)
intensity×horses	0.007(0.004)
intensity×pigs	−0.0001(0.0004)
intensity×goats	0.0001(0.0003)
intensity×self-consumptions	0.0001(0.0003)
intensity×cows sold	−0.001 (0.001)
intensity×cows bought	−0.001(0.001)
intensity×pigs sold	−0.00001(0.001)
intensity×pigs bought	0.001 (0.001)
intensity×goats sold	−0.0001(0.001)
intensity×goats bought	0.00005 (0.001)
intensity×gifts	−0.001(0.0005)
intensity×thefts	0.0001(0.0001)
intensity×type of plot farming	0.003 (0.002)
intensity×npk (kg)	0.00000(0.00003)
intensity×urea (kg)	−0.0001(0.0001)
intensity×phosphates (kg)	−0.0001(0.0002)
intensity×solid pesticides (kg)	−0.00001(0.00001)
intensity×liquid pesticides (cl)	0.00000(0.00000)
intensity×herbicide (g)	0.00000 (0.00000)
intensity×herbicide (cl)	0.00001 (0.00001)
intensity×fungicide (g)	−0.00004(0.00005)
intensity×fungicide (cl)	−0.00001 (0.00001)
intensity×rodenticide (g)	0.00000(0.00001)
intensity×rodenticide (cl)	0.00002(0.00002)
intensity×number of household members	0.0001(0.001)
intensity×average age	−0.0001 (0.0002)
intensity×number of families in the household	−0.0001(0.006)
intensity×number of children	0.001(0.001)
intensity×number of young working	−0.0003 (0.001)
intensity×number of women working	−0.0003(0.001)
Observations	51,398
R^2	0.336
Adjusted R^2	0.290
Residual Std. Error	1.209 (df = 48072)
Time and household f.e., village clustering	*p<0.1; **p<0.05; ***p<0.01

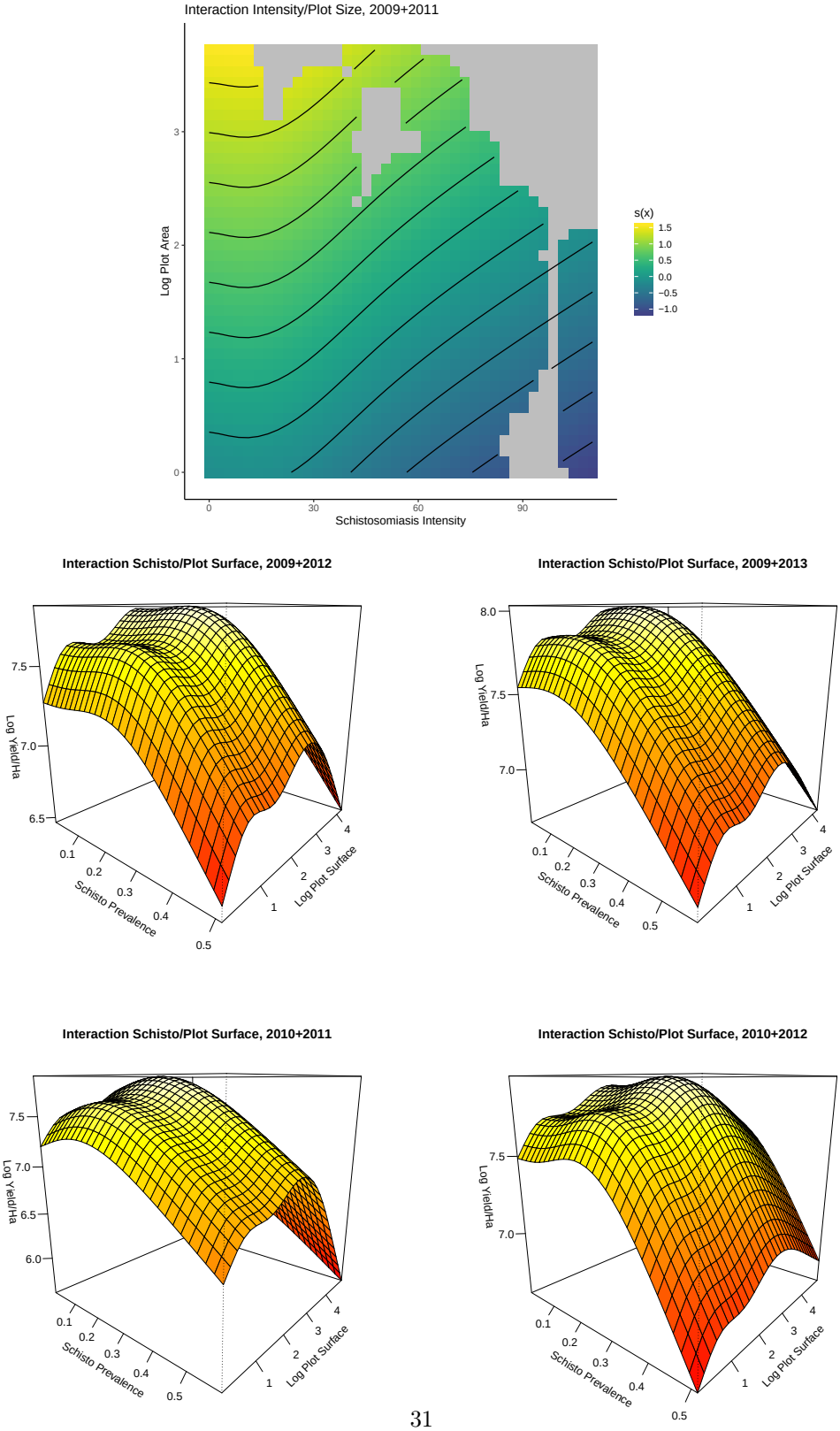
E Other determinants of agriculture in the instrumented linear regressions

Table 7 reports the rest of the coefficients for the linear estimations. For clarity we only report the ones significant at standard levels of confidence: because of the nature of the data, constant at a village level, and the procedure being 2SLS, errors are cluster-bootstrapped at a village level, and therefore increase largely estimation errors and many controls result statistically insignificant. All continuous covariates are in logs. Signs are usually as expected, with an interesting negative quadratic relationship between plot surface and yield, suggesting decreasing returns to scale in plot size.

Table 7 Other determinants of agricultural yield

	Dependent variable: log yield
Log surface	0.466^{***}
	(0.054)
Log surface^2	-0.140^{***}
	(0.025)
Presence of hired labor	0.106^{***}
	(0.033)
Total livestock used in agriculture	0.057^{***}
	(0.022)
Female livestock used for breeding	-0.051^{**}
	(0.024)
Cows	0.029^{*}
	(0.015)
Sales of cows	-0.049^{*}
	(0.026)
Npk (kg)	0.020^{**}
	(0.008)
Urea (kg)	0.028^{***}
	(0.009)
Solid pesticides (g)	0.027^{*}
	(0.014)
Herbicides (g)	0.015^{*}
	(0.008)
Plot near houses	0.604^{**}
	(0.283)
Plot near bushes	0.577^{**}
	(0.283)
Plot near encampment	0.587^{**}
	(0.287)
Plot on flatland	0.185^{***}
	(0.040)
Plot farmed with half-moon technique	-0.665^{***}
	(0.203)
Dead/live hedges	0.169^{**}
	(0.080)
Maximum level of education in household: Non-alphabetized	-0.123^{**}
	(0.057)
Maximum level of education in household: Alphabetized	-0.057^{*}
	(0.034)
Maximum level of education in household: Primary	-0.162^{**}
	(0.073)
Loss due to flooding	-0.982^{***}
	(0.122)
Loss due to fire	-0.549^{***}
	(0.117)
Loss due to drought	-0.680^{***}
	(0.044)
Loss after harvesting	-0.987^{***}
	(0.156)
Mean temperature	-0.04^{*}
	(2.597)
Temperature in dry season	0.03^{**}
	(1.586)
Precipitation in dry season	0.721^{***}
	(0.227)
Max dry spell	-0.307^{***}
	(0.113)
Mean of land surface temperature (day)	-0.02^{**}
	(1.088)
Mean of land surface temperature (night)	0.022^{***}
	(0.915)
Enhanced Vegetation Index (level)	4.8^{**}
	(2.014)
Note: Time+hh+crop fe. Cluster-bootstrapped (village)	*p<0.1; **p<0.05; ***p<0.01

Figure 12 (Upper panel) Interaction of schistosomiasis intensity and plot surface (Other panels) Yield and plot surface for different levels of disease intensity



F Schistosomiasis and cultivated plot surface

The lower four panels of 12 shows nonlinear estimates of the joint effect of schistosomiasis and the size of the cultivated plot. The negative quadratic relationship between yield and plot size remains established, and it seems that middle-sized plots are the ones that generate the higher yield. Regarding the *interacted* effect of schistosomiasis and plot surface, which describes the actual loss of yield due to schistosomiasis, for the years 2009-2011 the figure in level sets shows that all data for households in high infections cultivate smaller plots, and large plots suffer a lesser loss. This estimation is shown in the upper panel of Figure 12.

Table 8 Estimates of median regression

	Value	Std. Error	Pr(>|t|)
(Intercept)	0.513	0.004	0
Log Surface	0.291	0.018	0
Log Surface^2	−0.083	0.005	0
Presence of Hired Labor	0.040	0.016	0.011
Presence of Labor from Mutual Aid	0.042	0.011	0.0001
Collective Plot Management	0.008	0.009	0.387
Presence of Manual Labor	0.001	0.043	0.985
Gifts	−0.011	0.006	0.068
Male livestock for cattle raising	0.016	0.009	0.070
Female livestock for cattle raising	−0.024	0.012	0.043
Livestock Deaths	−0.011	0.005	0.022
Ovines	−0.029	0.007	0.0001
Sales of Ovines	0.024	0.008	0.003
Length of set-aside	−0.024	0.008	0.003
Npk (kg)	0.011	0.004	0.006
Urea (kg)	0.021	0.005	0.00004
Herbicide (g)	0.008	0.003	0.027
Herbicide (cl)	0.009	0.003	0.001
Rodenticide (g)	0.015	0.004	0.0004
Plot is on flatland	0.067	0.016	0.00003
Plot has a Filtering Dam	−0.113	0.074	0.128
Plot has half-moon	0.137	0.031	0.00001
Plot has soil banks	0.087	0.038	0.020
Plot has been loaned	−0.170	0.069	0.014
Plot has been inherited	−0.159	0.078	0.042
Plot has land holding rights	0.220	0.134	0.100
Plot is regularly owned	0.215	0.083	0.010
Presence of labor with yoked cows	0.026	0.013	0.043
Presence of motorized labor	0.029	0.012	0.018
Household is not literate	−0.081	0.025	0.001
Maximum household education is *École rurale*	−0.079	0.026	0.003
Loss due to flooding	−0.580	0.020	0
Loss due to parasites	−0.375	0.043	0
Loss due to dryness	−0.442	0.010	0
Post-harvest losses	−0.241	0.023	0
Number of young adults	0.016	0.007	0.017
Household does other rainfed cultures	0.170	0.078	0.029
Household does arboriculture	0.090	0.020	0.00001
Day temperature	−0.217	0.023	0
Night temperature	0.704	0.094	0
Precipitation in wet season	−0.209	0.065	0.001
Temperature in dry season	0.373	0.051	0
Max dry spell	−0.134	0.029	0.00000
Day land surface temperature year mean	−4.474	0.568	0
EVI year mean (level)	8.253	1.146	0
NDVI year mean (level)	−2.446	0.793	0.002

Note: only estimates significant at max 10% confidence level reported