# The Economic Impact of Schistosomiasis

Daniele Rinaldo [1]

with

Javier Perez-Saez [2]   Penelope Vounatsou [3]   Jean-Louis Arcand [1]

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

March 2, 2020

# Schistosomiasis (Snail Fever / Bilharzia)

# Schistosomiasis (2)

- Chronic and debilitating neglected tropical disease
- Water-based, with a complex lifecycle
- Parasitical disease for which snails are intermediate host
- Humans get infected through water contact (uro-fecal transmission)

Patterns of disease burden / transmission

↓

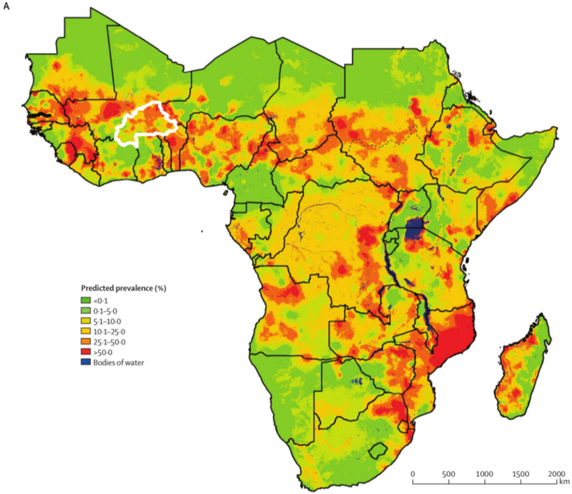
**anthropic factors** such as water resources development and human mobility (Steinmann et al. 2006)

# Schistosomiasis Facts

- The second most prevalent tropical parasitic disease after malaria, affecting about 200-300 million people worldwide (WHO, 2017)
- Loss of 3 million DALY/yr in the past decade, 190 million people still infected in 2016 (Global Burden of Disease Collaborative Network, 2016)
- Lifelong health effects: kidney failure, heightened HIV risk, bladder cancer, liver fibrosis
- Highest burden is on children: stunted growth, anemia, cognitive deficits and school drop-outs (King, 2010)

**Labour-impairing and poverty-reinforcing**

# Schistosomiasis Endemicity and Africa




(Lai et al., 2015)

# Research Question

- What we already know:
  - Link agricultural production / growth in developing countries (starting from World Bank, 1982)
  - Agriculture as tool against poverty (Christiaensen, Demery and Kühl, 2006, Audibert, 2011)
  - Schistosomiasis and water resources development (Steinmann et al. 2006, Perez-Saez et al., 2016)

## What we want to find:

→ **Impact of schistosomiasis on agricultural production**, and consequently on the development of countries in which it is endemic

→ Overall aim: explore the interlinked dynamics of **economic development, disease diffusion and water resources development**

# Burkina Faso

- Low-income, landlocked Sub-Saharan country, limited resources (World Bank 2017).
- Ca. 18.6 million inhabitants in 2016
- Main component of its economy is agriculture ($\sim$80% of the active population employed)
  - $\rightarrow$ Mostly subsistence production with low crop and livestock productivity (World Bank 2017)
  - $\rightarrow$ Low diversification, although increasing.
  - $\rightarrow$ **Cotton**: most important cash crop
- Burkina Faso is one of the countries in which schistosomiasis is endemic (Poda et al., 2004)

# Why Burkina Faso?

**Baseline case for worst possible burden**

→ higher aggregation: more confounding effects

→ constraints expected to be more binding: less substitution effect vs. an economy with more developed tertiary sectors

# Literature on Disease and Development

- Commission on Macroeconomics and Health (WHO, 2001) concluded that diseases are a barrier to economic growth.
- Several studies on how negative health shocks in child development lead to decreases in future wages or labour productivity (Sahn and Alderman (1998); Strauss and Thomas, 1998; Behrman and Rosenzweig, 2004)
- Well-known macro paper on health and development: Acemoglu and Johnson (2006)
  - → controversial: no particular evidence of increase in life expectancy leading to faster growth
- More recent work shows how the the links between health and economic development are more complex: Bloom, Canning and Sevilla, 2004; Bloom and Sachs, 1998; Gallup and Sachs, 2001; Barro, 1997; Kloos et al. 2008; King, 2010.

# Literature on Endemic Diseases and Development

- Lots of work in '60s - 80's for the effect of endemic/parasitic diseases on agriculture:
  - → on production (Foster, 1967; Barbosa and da Costa, 1981; Audibert, 1986): no clear effect of malaria, small effect of schistosomiasis
  - → on productivity (Baldwin and Weisbrod, 1974): only few significant effects
  - → reallocation of labour and land (Conly, 1975): no effect of both malaria and schistosomiasis
- More recent work:
  - → Malaria and agriculture: Audibert et al., 2003a, 2003b, 2011 (field data), mixed results
  - → Schistosomiasis and agriculture: Audibert and Etard (2008), no direct effect but impacts household behaviour

# Data

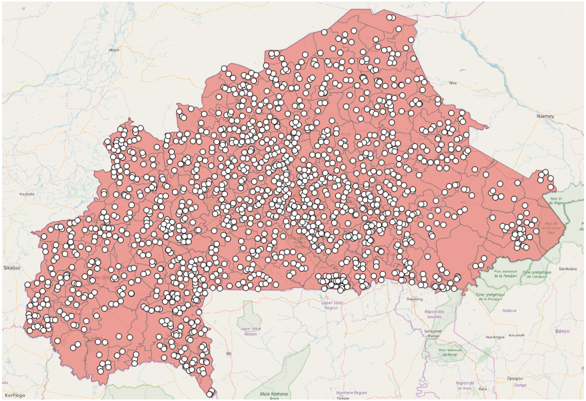
- **Enquétes Permanentes Agricoles (EPA)**, collected by the INSD (Institut National de la Statistique et de la Démographie), Ouagadougou

→ Large agricultural production data, 2003-2017
  - Plot-level data with information on crops, yields, inputs and pesticides, plot characteristics and labour type
  - Also information on contractual characteristics (sharecropping/owner-operator)

→ Household survey data, 1996-2017
  - Data on household characteristics and demographics
  - Rich data on livestock type and use
  - We merge and syncronize all data, then geo-localize 2585 villages

# Data (2)

- We merge and synchronize all data, then geo-localize 1,950 villages
  - → fuzzy string matching for names (we lost ca. 20)
  - → synchronized with administrative geo-spatial data
  - villages, 19,993 unique households, each on average cultivating around 10 plots
  - 10343 households observed at least 3 years throughout the dataset



# Climate Data: Remote Sensing Data Sources (Satellite)

→ Temperature and rainfall data: Google Earth + NOAA ESRL Physical Sciences Division (PSD)

→ Land Surface Temperature (LST) day/night: Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra

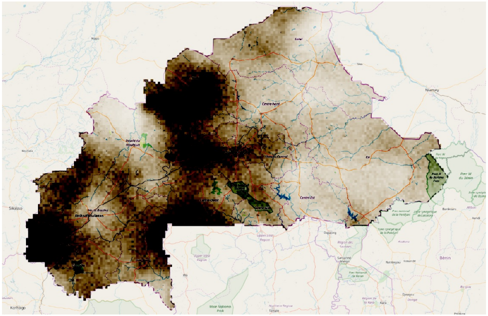
→ Normalized difference vegetation index (NDVI): MODIS/Terra

→ Enhanced vegetation index (EVI): STPH

- We create numerous controls from raw collected rasters: max/mean dry spells in dry/wet season, precipitation in dry/wet season, temperature shifts....

# Disease Data

- Bayesian geostatistical logistic model fitting with Bayesian variable selection. We have **two**: before and after 2010
- Model-based, high-resolution estimates of the prevalence of *Schistosoma* infection in school-aged children (median):




→ similarly, we include the **malaria** prediction median grid

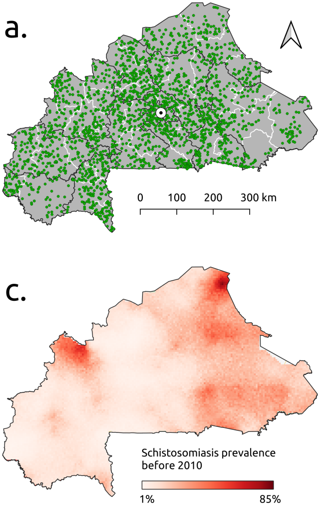
# Water Resources Data

- First focus is on four main dams:
  - Bagré Dam in Boulgou province, which also involves Zoundwéogo, Kouritenga and Ganzourgou provinces
  - Kompienga Dam in Kompienga province
  - Ziga Dam in Oubritenga province
  - Lery Dam in Nayala province

  → have irrigation purposes but mostly hydroelectrical

  → aimed at cities / larger agglomerates
- We then calculate the distance of *each village* from both large dams and medium/small dams and reservoirs

# All Data Together



# Dataset Creation

- Two disease data points per pixel: have to choose **two years**
  - $\rightarrow$ raster resolution makes climate+disease data **constant** at a village level
- Estimations with all years yield reasonable results but lots of potential issues
- Choose 2009 and 2011 to max obs + have a balanced panel
  - $\rightarrow$ also for model diagnostics
  - $\rightarrow$ all results hold for other combinations with enough sample size

# Schistosomiasis Intensity

- From prevalence to **intensity of disease** (av. worms/person)
- Survey data on ca. 200 points with both intensity and prevalence $\rightarrow$ fit (MLE) a negative binomial distribution

$$\text{Prev} = 1 - \left(1 + \frac{\text{int}}{k(\text{int})}\right)^{k(\text{int})} \tag{1}$$

for three different forms of $k$: constant, linear or quadratic

$\rightarrow$ AIC criterion: we choose a linear form $k = a \times \text{int} + b$, for which $\hat{a} = 0.01014$, $\hat{b} = 0.00112$.

$\rightarrow$ zeroes of (1) yield the intensity from prevalence map

# Model

- We posit the effect of schistosomiasis as a *pure productivity shock*:

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

  - $\rightarrow$ $Y_{ihjt}$: yield (output/ha) of plot $i$ at time $t$ by household $h$ in village $j$
  - $\rightarrow$ $\phi_{hjt}$ is the effect of the disease calibrated by $\theta$
  - $\rightarrow$ $F(X_{ihjt})$ is the production technology (nested)
  - $\rightarrow$ $e_{ihjt} = \exp(\alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt})$ represents plot-level unobservables, decomposed into crop, household and time-specific components.

- Log-linearized model is quasilinear:

$$y_{ihjt} = \tilde{A}_{jh} - \theta\phi_{hjt} + \tilde{F}(X_{ihjt}) + \alpha_t + \alpha_h + \alpha_c + \epsilon_{ihjt}$$

- Use **intensity** $I_{jt}$ since $\phi$ is not observable. Goal: $\theta$

# Reduced form? Not really

- Is this compatible with household optimization? Depends
- We decompose the production in $\tilde{F}(X_{ihjt}, K_{ihjt})$, where the $k$ inputs $X$ are to be optimized
- Profit-maximizing conditions will impose $\tilde{F}(X^*_{ihjt}, K_{ihjt})$, where $X^*_{ihjt}(A_{hj}, I^{-\theta}_{jt}, p_{jt}, K_{ihjt}, \epsilon_{ihjt})$ is input use at the optimum.
  - $\rightarrow$ $p_{jt}$: input prices, $K_{ihjt}$: non-optimized inputs (climate)
- Then:

$$\frac{\partial y^*}{\partial I} = -\theta + \sum_{i=1}^{k} \frac{\partial \tilde{F}(X^*_i, K)}{\partial X_i} \frac{dX^*_i}{dI}$$

- If $\frac{dX^*_i}{dI} = 0$ for all $i = 1 \ldots k$ then the original specification is compatible with optima

## Quasilinear model as a model of effective labor supply

- Disease as shock to *effective* labor supply, power form for labor (can be generalized): :

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

- Maximization program for $L$ and the matrix $X = [X_1, \ldots X_k]$

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

- yields the optima $L^*(A, \phi^{-\tilde{\theta}}, K, w, p), X_i^*(A, \phi^{-\tilde{\theta}}, K, w, p)$, where $p$ is the vector of prices and $w$ is paid wages. Then, setting $\theta = \tilde{\theta}\alpha$ :

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

# So?

- Factor regressions for each of the inputs potentially subject to choice
  - → pesticides, labor variables, livestock (even household composition for good measure)
  - → also check all interactions with schistosomiasis intensity

**No significant signal detected for the disease**
**(i.e. all partial derivatives are estimated at zero)**

- Evidence of schistosomiasis as a shock to total factor productivity: compatible with the nature of its burden
- We then work on the quasilinear model

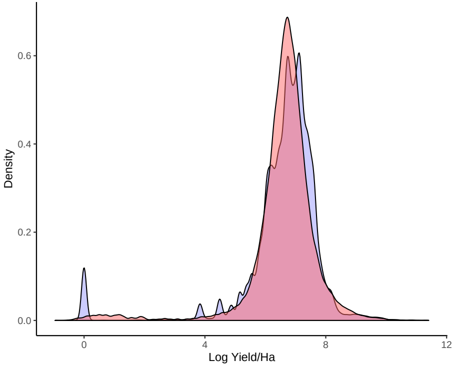
# Potential Issues

- Clustering
- Endogeneity/measurement error
- Shape of $F(X)$
- Spatial correlation

# Clustering + Aggregation

- Pixel-level raster resolution → Clustering at **villages** (estimates less precise)

→ density of plot yields vs. same density (rescaled) with hh fe partialled out:



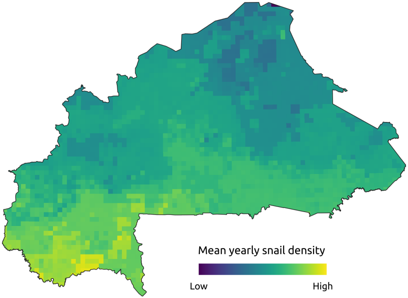

- Village aggregation: *representative household*

# Endogeneity/Measurement Error

- Model error for the disease intensity is accounted for
- But *effective* level of intensity for a household is likely to differ
  $\rightarrow$ measurement error
- Also garden variety endogeneity, though less a concern
- IV: **Intermediate host density**
  $\rightarrow$ Lewbel het-IV 2-step GMM procedure results in similar estimates

# Snails as IV

- Schistosomiasis: **two forms**, intestinal and urinary
  - → *S. mansoni*: southwest, snail host: *Biomphalaria pfeifferi*
  - → *S. haematobium*: overall, snail host: *Bulinus*
- Snail densities: **IV**
- Each village: density for each season (dry/rainy/winter), weighed on spatial prevalence of each species



# Production technology

- Cobb-Douglas $\rightarrow$ fully linear model
- Can we do better?

Interest is on identifying $\theta$!

$\downarrow$

**Adaptive methods** can be used to absorb the confounding effects given by the nonlinear function $\tilde{F}(X)$ from both $y$ and $I$

# Machine learning estimation

- We adapt to a new setting the methods of Belloni et al. (2017): **Double (Debiased) Machine Learning**
- **Double Neyman-orthogonal moment conditions**

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

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 for estimating $\tilde{F}$, and

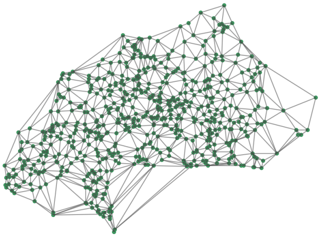
$$\eta = (\mathbb{E}[Y|X], \mathbb{E}[I|X])$$

- Empirical counterparts for $\widehat{\mathbb{E}[Y|X]}$ and $\widehat{\mathbb{E}[D|X]}$? **Machine learning estimators**

# Machine learning estimation (2)

- We choose four, and compare performance with MSE:
  - → **Lasso** with very (**very**) large control matrix (adapted for panel clustering)
  - → **Random Forests**
  - → **Neural Networks**
  - → **Gradient Boosting Machines**
- Intuition?
  - → 2nd moment condition valid under "local" mistakes in the nuisances.
  - → ML techniques are excellent for prediction/classification but obscure for interpretation/regression - that's ok here!
- Procedure requires **at least** 2-fold cross-validation in order for it to be unbiased

# Spatial Correlation



- Adjacency effects $\rightarrow$ **village distances** $d$ as spatial weights
- Spatial autoregressive model ($W$: matrix of weights):

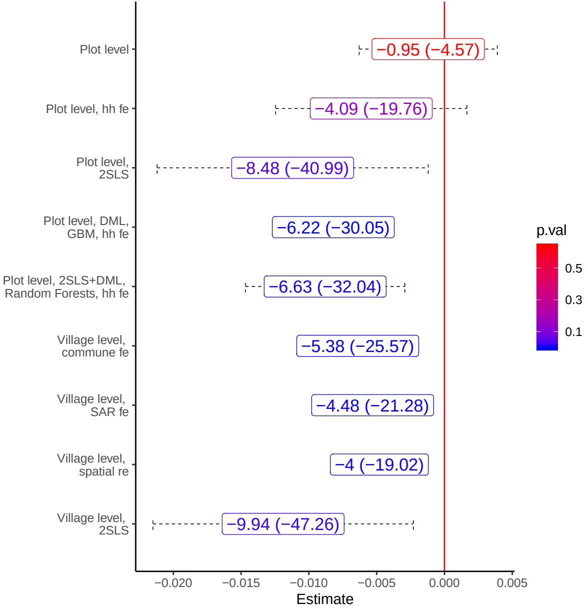
$$y_{jt} = (1_{2j} - \rho W_j)^{-1}[I_{jt}\theta + \tilde{X}_{jt}\beta + \alpha_t + \epsilon_{jt}]$$

- Spatial random effects model:

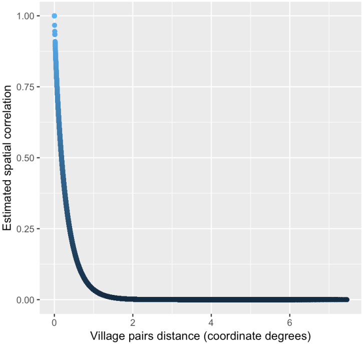
$$\begin{aligned} y_{jt} &\sim N(I_{jt}\theta + \tilde{X}_{jt}\beta + \alpha_t, C_\nu(d)) \\ C_\nu(d) &= \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{d}{\rho}\right) \end{aligned}$$

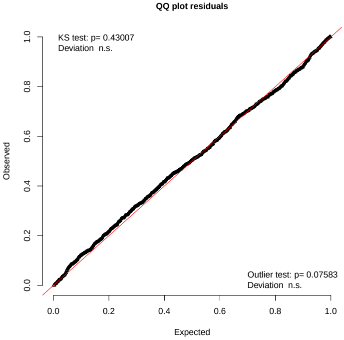
$d$: village pair distance, $\rho$: scaling, $\nu$: smoothing

# Results


Plot level
Plot level, hh fe
Plot level, 2SLS
Plot level, DML, GBM, hh fe
Plot level, 2SLS+DML, Random Forests, hh fe
Village level, commune fe
Village level, SAR fe
Village level, spatial re
Village level, 2SLS
−0.95 (−4.57)
−4.09 (−19.76)
−8.48 (−40.99)
−6.22 (−30.05)
−6.63 (−32.04)
−5.38 (−25.57)
−4.48 (−21.28)
−4 (−19.02)
−9.94 (−47.26)
−0.020
−0.015
−0.010
−0.005
0.000
0.005
Estimate
p.val
0.5
0.3
0.1

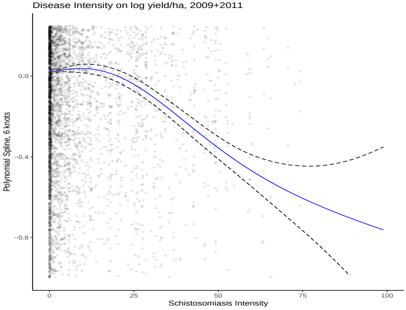
# Spatial effects + model fit

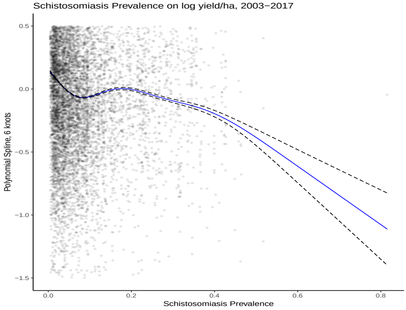


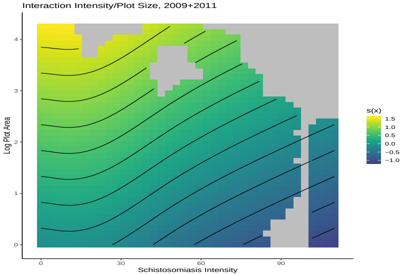



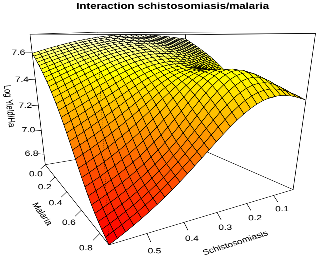

- Moran's I goes from 0.13 (significant) to 0.07 (non-significant) by only taking commune fixed effects

# Non-parametric estimates

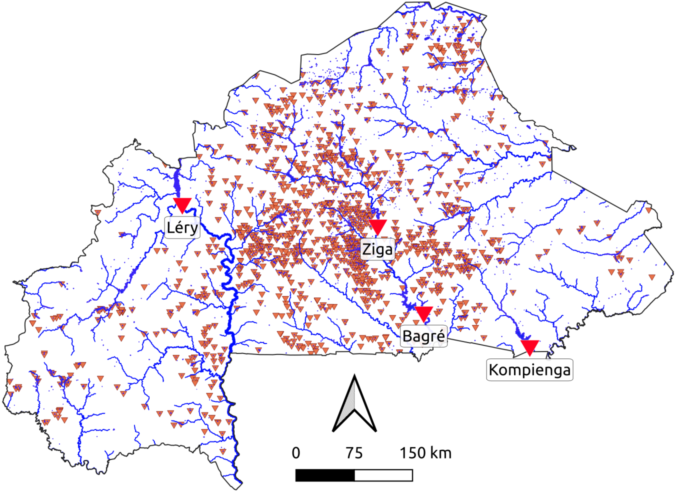


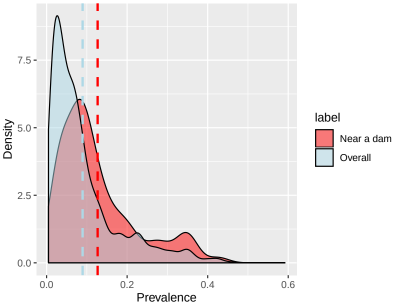



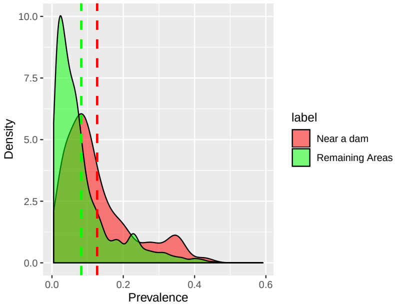
# Schistosomiasis, Water Resources and Development



# Schistosomiasis, Water Resources and Development (2)

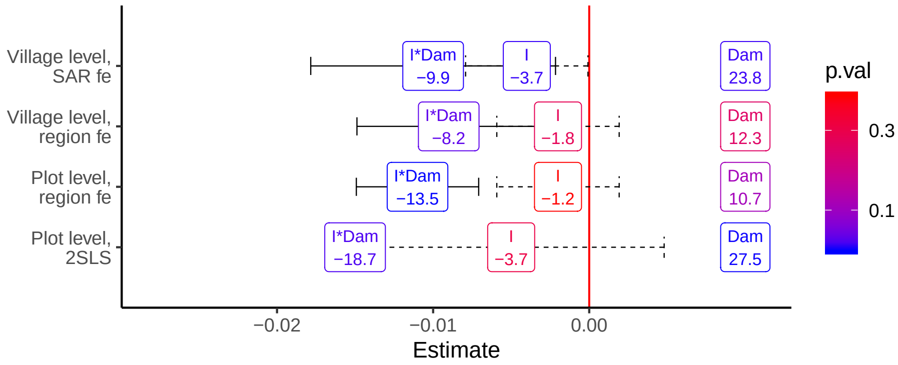
- First pass:
  - $\rightarrow$ Impact of four main dams $I_{jt} \times \mathbb{1}_{dam}$



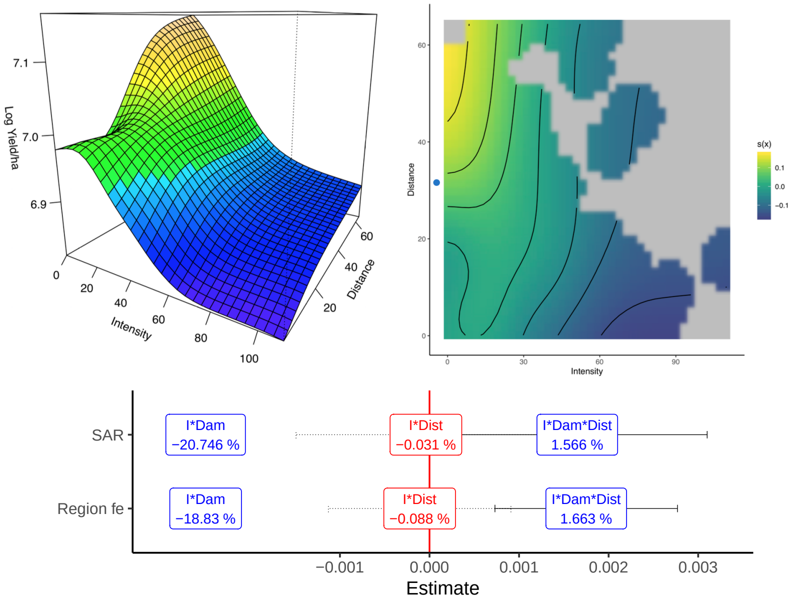




- Then:
  - $\rightarrow$ Impact of distance from **all** dams/irrigation reservoirs: $I_{jt} \times dist_{jt}$
  - $\rightarrow$ Impact of distance from large dams: $I_{jt} \times \mathbb{1}_{dam} \times dist_{jt}$

# Schistosomiasis, Water Resources and Development (3)




→ spatial effects now matter! Moran's I for model with region fe = 0.1, p.val $2e^{-5}$


Log Yield/ha
7.1
7.0
6.9
0
20
40
60
80
100
Intensity
20
40
60
Distance
s(x)
0.1
0.0
−0.1
Distance
60
40
20
0
0
30
60
90
Intensity
SAR
Region fe
I*Dam
−20.746 %
I*Dist
−0.031 %
I*Dam*Dist
1.566 %
I*Dam
−18.83 %
I*Dist
−0.088 %
I*Dam*Dist
1.663 %
−0.001
0.000
0.001
0.002
0.003
Estimate

# Implications and policy relevance

- Inequality: most affected households/villages are not those who reap the benefits of large infrastructures
- Productivity shock $\rightarrow$ Poverty-reinforcing, no apparent substitution effects
- Policy-wise:
  - $\rightarrow$ Largest economic burden is at upper 25% quantiles of intensity
  - $\rightarrow$ Reinforcement of small- to middle-sized irrigation networks

# Conclusions

- We use the case of Burkina Faso to establish a baseline case for the economic impact of schistosomiasis
- Large, negative, causal and nonlinear effect
- Complex interactions with climate and malaria
- Compatible with a household decision framework
  - $\rightarrow$ Effect on production as a productivity shock
- Feedback effect of water resources development

# Thank you!