economies

Article

Revisiting the Impact of Dams on Malaria and Agriculture

Sebastien Mary [1,*], Kyle Craven [2], Avraham Stoler [3] and Sarah Shafiq [3]

[1] Department of Economics, Governors State University, University Park, IL 60484, USA
[2] Chicago Public Schools, Chicago, IL 60602, USA
[3] Department of Economics, DePaul University, Chicago, IL 60614, USA
* Correspondence: smary@govst.edu

Abstract: We estimate the effect of large dams on malaria incidence in India between 1975 and 1995. We combine instrumental variables approach with a panel model with unobserved common factors allowing us to fully capture the endogeneity of dam location and unobserved time-varying heterogeneity. Dams result in increased malaria incidence in districts where dams are located and in downstream areas. We find that the construction of a large dam increases a district's annual malaria incidence by about 0.9 to 1.4 percent, and by about 1 to 1.5 percent in downstream districts. We also find that this malaria-increasing effect of dams persists over time. Our results imply that the construction of dams in malaria-sensitive areas should be coupled with direct interventions, such as the wide deployment of insecticide-treated nets or the roll-out of future vaccines. Furthermore, we examine the contribution of agricultural development to this malaria-increasing effect of dams. We find that dam construction benefits agriculture in the vicinity of dams, as well as in downstream areas. These positive effects are driven by increased irrigation and cultivation in the vicinity of dams, while they are driven by changes in cropping patterns in downstream areas, where the cultivation of high-yielding variety crops increases. Finally, a back-of-the-envelope calculation suggests that the agricultural production gains from dam construction dominate the economic losses resulting from increased malaria.

Keywords: irrigation; large dams; malaria; agriculture; India; SDG 3.3

JEL Classification: Q19; O20

Citation: Mary, Sebastien, Kyle Craven, Avraham Stoler, and Sarah Shafiq. 2023. Revisiting the Impact of Dams on Malaria and Agriculture. *Economies* 11: 173. https://doi.org/10.3390/economies11070173

Academic Editor: Ralf Fendel

Received: 5 May 2023
Revised: 18 June 2023
Accepted: 19 June 2023
Published: 25 June 2023

1. Introduction

Large irrigation dams play a vital role in the implementation of several key Sustainable Development Goals (SDG), because they potentially provide a range of benefits, from preventing river floods, improving access to electricity and reliable clean water, strengthening the resilience of agricultural systems in the face of climate change, reducing poverty and food insecurity, to promoting economic and social development (e.g., Duflo and Pande 2007b; Kibret 2018; Dillon and Fishman 2019).

However, large dam construction has been met with strong opposition from economists and environmentalist groups. For instance, dams are criticized for contributing to local inequality because (most of) the gains associated with dams may largely benefit the command area, typically downstream, while the costs might be concentrated in areas where dams are located, or the catchment area (e.g., Singh 1990; Duflo and Pande 2007b). Dams might come with a range of negative environmental and economic impacts, e.g., increased silting, salinity, or pollution risks (e.g., Mettetal 2019), because they fragment rivers and disrupt their flows, eventually threatening ecosystems that people and wildlife depend on for their survival (McCartney and Sally 2007; FitzHugh and Vogel 2011; Wildi 2010).

Furthermore, research into the health impacts of large dams has also heightened through recent case studies and anecdotal evidence (e.g., Dillon and Fishman 2019). This line of research particularly focuses on the impact that large dams have on food and water scarcity, communicable diseases, social disruption, and vector-borne diseases such as schistosomiasis and malaria (e.g., Lerer and Scudder 1999; Kibret 2018; Mettetal 2019). In particular, an increasing body of anecdotal reports and case studies suggest large dams are associated with increased malaria cases and deaths (e.g., Yewhalaw et al. 2013; Dejenie et al. 2012). This paper focuses on the causal link between large dams and malaria incidence.

Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected female Anopheles mosquitoes. Once constructed, large dams can create small pools of standing water, which is an ideal breeding and egg-laying ground for adult female Anopheles mosquitoes, the primary vector of malaria transmission (International Water Management Institute 2018; Sadoff 2018). These small pools of water where predators to mosquitoes may be absent (e.g., fish or tadpoles) (e.g., Kibret 2018), often caused by flooding within the vicinity and upstream of the dam, create a year-round habitat for Anopheles mosquitoes, extending the period of malaria transmission in some regions to include both the rainy and dry seasons (Mbakop et al. 2019). Furthermore, the slope of the dam can also affect the Anopheles population in the area. In fact, dams with gentle and shallow slopes allow pools of water to easily form, which further increases the probability of Anopheles habitat formation.

Conversely, large dams can disrupt the formation of Anopheles breeding habitats that are located downstream of the dam by reducing river flows and decreasing the amount of water discharge in the area (Lee et al. 2019; Larinier 2001). This reduction in the water volume downstream of the dam limits the possibility of Anopheles habitat formation, which can greatly reduce the transmission of malaria in the area. This effect is especially prominent during the dry season.

Large dams can also exacerbate the impact that agricultural development has on malaria transmission in a variety of ways (e.g., Janko et al. 2018). First, dams can greatly expand the role of irrigation as an agricultural practice in nearby farmlands. If these irrigation systems lack efficient drainage, water can accumulate into small standing pools, providing additional breeding grounds for Anopheles mosquitoes (GBCHealth 2012). Other mechanisms by which dams can facilitate an increase in malaria incidence through agriculture expansion include deforestation, increased crop cover, and agriculture encroachment in highland regions (Asenso-Okyere et al. 2009; MacDonald and Mordecai 2010; Guerra et al. 2006). Finally, migrants from malaria endemic areas seeking work or food and water security can travel to these newly developed farmlands, which could also greatly facilitate the transmission of malaria (Falavigna-Guilherme et al. 2005).

While more than 200 million people are affected by malaria worldwide[1], the World Health Organization aims at reducing global malaria cases and mortality rates by at least 90 percent by 2030. The push to end malaria is also recognized in the Sustainable Development Goals Target 3.3 "*By 2030, end the epidemics of AIDS, tuberculosis, malaria and neglected tropical diseases and combat hepatitis, water-borne diseases and other communicable diseases*". Given this policy background, this paper examines the impact of large dams on malaria incidence.

Against this conceptual background, the empirical literature has mainly consisted of case studies and anecdotal evidence that come up with mixed conclusions. For example, Lautze et al. (2007) analyze the impact that the Koka dam had on malaria incidence in the Rift Valley of Ethiopia from 1994 to 2002. They find that malaria case rates among people living within 3 km of the Koka reservoir were 1.5 times greater than those living between 3 and 6 km from the reservoir and 2.3 times greater than those living 6–9 km from the reservoir. Additionally, Ghebreyesus et al. (1999) find that the incidence of malaria is significantly higher in children who live in villages close to dams (within 3 km) than in children who live further away (between 8–10 km). Recently, Kibret et al. (2019a) found that distance to reservoir shoreline and malaria incidence were negatively correlated in their analysis of three dams in Ethiopia[2] from 2010–2014.

In contrast, Yewhalaw et al. (2013) study the impacts that an Ethiopian dam, the Gilgel-Gibe dam, had on malaria incidence from 2008 to 2010 but found no statistically significant relationship between malaria incidence and proximity to the dam. However, the authors do find that the malaria vector (Anopheles arbiensis) was more abundant in villages closer to the reservoir. Likewise, in their study of six dams in Ethiopia's Tigray Region, Dejenie et al. (2012) find that malaria prevalence did not increase in villages within 4 km of a dam compared to villages further away[3].

Ndiath et al. (2012) study the impact of agricultural development on malaria transmission in the Senegal River basin after the construction of two dams along the Senegal River. They find that malaria transmission was low and seasonal during their two-year study period. Interestingly, the authors also find that Anopheles densities were significantly correlated with the presence of ditch water (for activities like animal watering and brick manufacturing) but not correlated with proximity to the river. Furthermore, Janko et al. (2018) find that rural children under the age of 5 in the Democratic Republic of Congo were particularly vulnerable to malaria transmission resulting from similar agricultural development. Falavigna-Guilherme et al. (2005) find that the construction of the Itaipu Dam on the Paraná River resulted in the proliferation of the Anopheles mosquito in the area and the subsequent malaria outbreak.

In their extensive analysis of the impact dams have on malaria incidence in sub-Saharan Africa, Kibret et al. (2019b) suggest that between 0.7 and 1.6 million malaria cases per year were attributable to large dams from 2000–2015. Further, in their univariate correlation analysis, the authors find that the slope of the reservoir explained 46.8 percent of the variation in malaria incidence around large dams, more than any other environmental factors, including rainfall, temperature, and humidity. They also find that the slope of a reservoir significantly impacted the malaria incidence in communities less than 5 km from a dam.

An important limitation of the literature is that existing studies remain descriptive and vulnerable to several biases, among which is the endogeneity bias due to the non-random nature of dam location, and thus only provide non-causal evidence (e.g., Duflo and Pande 2007b; Dillon and Fishman 2019)[4]. To establish potential causality between dams and malaria is indeed challenging from an econometric standpoint, and virtually all existing studies have failed to do so.

A notable exception is Duflo and Pande (2007b)—DP thereafter—who provide a viable causal identification strategy by relying on various geographical features of the district. To be more specific, they exploit variation in dam construction induced by differences in river gradient across districts within Indian states to construct instrumental variable estimates and find no effect of large dams on malaria incidence in India.

Overall, due to the regional focus, the somewhat mixed results, and existing methodological limitations in the literature, there is little convincing evidence that dams cause malaria incidence. In fact, extensive systematic reviews of the literature also confirm the lack of evidence (e.g., Keiser et al. 2005). For example, Keiser et al. (2005) conclude that "no clear relationship emerges from this literature review between the risk factor "development and operation of water projects" [sic] and malaria risk". Dillon and Fishman (2019) suggest that "the pathway of higher malaria incidence related to increased hydrological investments is not well substantiated". These reviews also highlight the dearth of evidence in the empirical literature and call for additional research[5].

While DP's identification strategy deals with the endogeneity bias, they rely on state-year fixed effects to partially account for time-varying omitted variables. Doing so imposes that the impact of unobservable common shocks is homogenous on malaria across all districts within a state. This is unlikely to be valid for several reasons because state-level common shocks, for example, climate shocks or input/output price shocks, would likely have differentiated effects on malaria incidence across districts within a state, dependent on the district's topology, type of agricultural practices and systems, health systems and infrastructures, and existing mitigation and adaptation capacities (e.g., Skoufias 2003; Kuriakose et al. 2013; Asfaw et al. 2017).

More generally, there is no reason to assume that the impact of such common shocks would not be heterogeneous across districts. Assuming otherwise may result in biased estimates. For example, Mary (2022) shows that restrictive identification assumptions with respect to modelling time dependence, including the use of state-year fixed effects, may create an omitted variable bias and lead to largely misguided policy implications in the case of the mitigating impact of dams on ethnic conflict in the face of rainfall shocks in India.

Recently, models with unobserved common factors or interactive fixed effects have been increasingly proposed to solve similar issues in the general economics literature. Such an approach captures time-varying heterogeneity via a set of factor loadings and requires the imposition of a factor structure. While common factors or interactive effects models are somewhat more complicated than fixed effects models, they have increasingly become commonplace in the empirical economics literature (e.g., Gobillon and Magnac 2016).

The literature has provided several modeling approaches to estimate panel models with unobserved common factors or interactive effects: common correlated effects (CCE) (Pesaran 2006), interactive principal components (IPC) (Bai 2009; Bai and Li 2014), and general two-stage IV (G2SIV) approaches (Norkutė et al. 2021; Cui et al. 2022). All these approaches have benefits and drawbacks. However, the latter may be a preferred option because it combines the attractive features of both the CCE and IPC approaches (Kripfganz and Sarafidis 2021).

In this paper, we build on Duflo and Pande (2007b) and combine their instrumental variables (IV) approach with the G2SIV approach for estimating panel models with unobserved common factors from Norkutė et al. (2021) and Cui et al. (2022) to control for, respectively, the endogeneity of dam placement, and time-varying unobserved heterogeneity. Given our modeling approach, we assess the effect of technologically feasible dams on malaria.

Our paper provides evidence that large irrigation dams increase malaria incidence. We find that dams increase malaria incidence in districts where dams are located and in downstream districts. Our estimates imply a one-dam increase in the number of dams in the district increases annual malaria incidence by about 0.9 to 1.4 percent on average. The effect in downstream districts is about 1 to 1.5 percent. This confirms the multiple anecdotal and descriptive reports that have suggested a potential link between dams and malaria (e.g., Kibret et al. 2019b; Lautze et al. 2007; Ghebreyesus et al. 1999).

To understand the mechanisms linking dams and malaria, we provide a second set of agriculture-focused results. These results suggest that dam construction benefits agriculture in the vicinity of dams, as well as in downstream areas. These positive effects are driven by increased irrigation and cultivation in the vicinity of dams, while they are driven by changes in cropping patterns in downstream areas. In particular, the cultivation of high-yielding variety (HYV) crops[6], which require more water, increases in downstream areas. These agricultural results highlight the potential mechanisms between dam construction, irrigation area, cropping patterns, and malaria incidence.

Overall, we contribute to the literature on irrigation dams in three important ways. First, from an empirical standpoint, we show that large dams affect malaria incidence in India. Second, we show that dam construction benefits agriculture in the vicinity of dams, as well as in downstream areas. These findings are novel in the empirical literature and will contribute to a re-assessment of the distributional cost-benefit analyses linked to the construction of large dam infrastructures from both health and economic perspectives. Third, from a methodological perspective, we highlight the role of time-varying heterogeneity in empirical studies examining the impact of irrigation infrastructures, the limitation of state-year fixed effects to account for unobserved time-varying heterogeneity in this context, and finally demonstrate the usefulness of common factors approaches.

The rest of the paper is structured as follows. Section 2 describes the identification strategy. Section 3 presents the data. Section 4 analyses the empirical results while Section 5 concludes.

2. Empirical Strategy

2.1. Empirical Model

To examine the impact of dams on malaria incidence, we can start with the following model that has been used in previous studies:

$$y_{ist} = \beta D_{ist} + \alpha D^{U}_{ist} + \gamma Z_{ist} + \delta Z^{U}_{ist} + \chi C_{ist} + \theta_i + \rho_{st} + \varepsilon_{ist} \quad (1)$$

where y_{ist} is the outcome of interest, i.e., malaria incidence, in district i in state s at time t; D_{ist} is the number of dams in the district; D^{U}_{ist} is the number of dams located upstream of the district; Z_{ist} and Z^{U}_{ist} is a vector of controls for the district and for upstream districts. In an extended model, C_{ist} is a vector of climatic variables, namely rainfall and temperature shocks. θ_i are district fixed effects, controlling for time-invariant characteristics. ρ_{st} are state-year (interacted) fixed effects partially controlling for time-varying heterogeneity with homogenous common shocks across districts in a state. β captures the impact of a dam on malaria in the district where the dam is located (i.e., the own district effect), while α describes the downstream impact of a dam.

There are two main problems with this model. First, the literature has recognized that dam construction may not be random. Indeed, local governments may target areas that are already more productive (but perhaps associated with higher malaria burden[7]) or politically better-connected (Mettetal 2019). By contrast, the construction of dams may be seen as an engine of growth in poorer districts.

To control for this endogeneity issue, several studies have used the IV strategy designed by Duflo and Pande (2007b)[8]. They exploit variation in dam construction induced by differences in river gradient across districts within states to construct IV estimates. Following their approach, we estimate:

$$D_{ist} = \alpha + \sum_{k=2}^{4} \varphi_k \left(RG_{ki} * \overline{D}_{st}\right) + \omega\left(M_i * \overline{D}_{st}\right) + \sum_{k=2}^{4} \pi_k \left(RG_{ki} * l_t\right) + \theta_i + \rho_{st} + \varepsilon_{ist} \quad (2)$$

where D_{ist} is the number of dams in the district; RG_{ki} are the river gradient variables; $\overline{D}_{st}$ is predicted dam incidence in the state; M_i is a vector of district-specific time invariant variables, i.e., district elevation, overall gradient measures, river length, and district area. Equation (2) also includes district fixed effects and state-year fixed effects. The interaction of river gradient variables with country-level common shocks $RG_{ki} * l_t$ accounts for national-time varying effects of river gradient on the outcomes of interest. The river gradients capture the proportion of the river with a gradient between 1.5 and 3%, 3 and 6%, and above 6%.

The parameters from Equation (2) can be used to predict the number of dams in the district, $\hat{D_{ist}}$, that is used as an instrument for D_{ist}. DP also rely on Equation (2) to predict the number of dams upstream, arguing that their approach is more efficient because it uses all available variation to predict the relationship between river geographic features and the number of dams (rather than only those upstream). If a district does not have an upstream district, $\hat{D^{U}_{ist}}$ is replaced with 0.

The next step of the empirical strategy comes down to estimating a marginally modified version of Equation (1):

$$y_{ist} = \beta \hat{D_{ist}} + \alpha \hat{D^{U}_{ist}} + \gamma Z_{ist} + \delta Z^{U}_{ist} + \chi C_{ist} + \theta_i + \rho_{st} + \varepsilon_{ist} \quad (3)$$

where Z_{ist} and Z^{U}_{ist} include, respectively, all controls from Equation (2) but the interacted river gradient variables with predicted dam incidence in the state. Estimating Equation (3) captures the local average treatment effect (LATE) of technologically feasible dams (or in other words, the LATE of dams constructed for geographical reasons).

The second problem, which has been much less discussed in the literature, relates to how unaccounted time-varying heterogeneity (TVH) may affect estimates (e.g., Blanc and Strobl 2014). To control for TVH, the literature related to dams has used a variety of fixes.

Duflo and Pande (2007b) use state-year fixed effects. Sarsons (2015) uses state-level time trends. Mettetal (2019) uses a combination of year fixed effects and district-level trends.

In the face of such diverse modeling choices, it is critical to understand their implications. Mary (2022) shows that restrictive assumptions with respect to modeling time dependence to partially capture time-varying heterogeneity may result in largely biased findings in the case of the links between rainfall shocks, dams, and ethnic conflict in India. Therefore, we argue that state-year fixed effects are not sufficient to fully address concerns about time-varying heterogeneity.

Ideally, one would like to replace state-year fixed effects with district-year fixed effects. This is of course in practice impossible due to the incidental parameter problem leading to biased and inconsistent estimates. To circumvent this challenge, common factor approaches[9] have been increasingly used to account for rich sources of unobserved heterogeneity. Following the introduction of common factor approaches, we can modify Equation (3):

$$y_{ist} = \beta \hat{D_{ist}} + \alpha \hat{D^U_{ist}} + \gamma Z_{ist} + \delta Z^U_{ist} + \boldsymbol{\chi} C_{ist} + \theta_i + \rho_{st} + \boldsymbol{\mu_i} F_{st} + \varepsilon_{ist} \tag{4}$$

where F_{st} is a set of common factors and μ_i a vector of factor loadings. μ_i represents the heterogeneous impact of unobservable common shocks within a state on each district. Let's finally note that estimating Equation (4) is exactly equivalent to estimating Equation (3) if common shocks have homogenous effects on the outcome of interest across districts within a state, $\boldsymbol{\mu_i} = \boldsymbol{\mu}$.

Furthermore, if we ignore the endogeneity of dam location, we estimate a modified version of Equation (4) where the numbers of dams in the district and upstream are not instrumented:

$$y_{ist} = \beta D_{ist} + \alpha D^U_{ist} + \gamma Z_{ist} + \delta Z^U_{ist} + \boldsymbol{\chi} C_{ist} + \theta_i + \rho_{st} + \boldsymbol{\mu_i} F_{st} + \varepsilon_{ist} \tag{5}$$

In other words, this latter model accounts for time-varying heterogeneity but ignores the non-random allocation of dams. The estimations of Equations (4) and (5), in essence, provide respectively IV- and OLS-like estimates in a common factors framework.

Last, we transform the models in Equations (4) and (5) by introducing a lagged dependent variable to account for the existence of persistence in malaria. We obtain, in the case of Equation (4):

$$y_{ist} = \sigma y_{ist-1} + \beta \hat{D_{ist}} + \alpha \hat{D^U_{ist}} + \gamma Z_{ist} + \delta Z^U_{ist} + \boldsymbol{\chi} C_{ist} + \theta_i + \rho_{st} + \boldsymbol{\mu_i} F_{st} + \varepsilon_{ist} \tag{6}$$

where σ captures persistence in malaria cases. In the case of Equation (5), we get:

$$y_{ist} = \sigma y_{ist-1} + \beta D_{ist} + \alpha D^U_{ist} + \gamma Z_{ist} + \delta Z^U_{ist} + \boldsymbol{\chi} C_{ist} + \theta_i + \rho_{st} + \boldsymbol{\mu_i} F_{st} + \varepsilon_{ist} \tag{7}$$

2.2. Estimation

The literature has provided several modeling approaches to estimate panel models with unobserved common factors or interactive effects: common correlated effects (CCE) (Pesaran 2006), interactive principal components (IPC) (Bai 2009; Bai and Li 2014), and general two-stage IV (G2SIV) approaches (Norkutė et al. 2021; Cui et al. 2022).

All these approaches have benefits and drawbacks. However, the latter may be a preferred option because it combines the attractive features of both the CCE and IPC approaches (Kripfganz and Sarafidis 2021). First, unlike CCE or IPC, the G2SIV approach does not suffer from incidental parameters bias and thus does not require bias correction. Second, the IV approach is robust and computationally inexpensive. Third, the approach allows for endogenous regressors (if external instruments are available). Fourth, it can be applied with unbalanced panels.

In the first stage, the IV approach[10] projects out the common factors from exogenous covariates using principal components analysis and constructs instruments from defactored covariates. In the second stage, the entire model is defactored based on factors extracted from the first-stage residuals, and IV regression is implemented again using the same instruments. This approach removes the common factors from the error term and the regressors separately in two stages.

There are a few practical choices to be made for the implementation of the IV approach. We mainly rely on the Ahn and Horenstein (2013) eigenvalue ratio test to compute the number of factors jointly. We set the maximum number of factors to 4 for each estimation stage and each set of instruments. Finally, note we double down on our common factors approach and also introduce common factors in Equation (2) to make the entire estimation strategy consistent. We also provide supporting evidence for why this is needed given the data.

3. Data

Data is mainly taken from the dataset compiled by (Duflo and Pande 2007a). The dataset we use includes information on district geography, dam placement, river gradient, and malaria incidence, between 1975 and 1995 for a sample of 358 Indian districts. While we refer the reader to DP for more details, we summarize key information about each variable below. Descriptive statistics can be found in Table A1 in Appendix A.

Data on dams come from the World Registry of Large Dams. A large dam is defined as a dam having a height of 15 m from the foundation, or, if the height is between 5 and 15 m, having a reservoir capacity of more than 3 million cubic meters. The registry lists the nearest city to the dam and allows for allocating dam to district.

The malaria incidence data used was collected by India's National Malaria Eradication Program (NMEP). Malaria incidence is measured as the log of annual parasite incidence (API), where API is defined as the number of smears positive for *P. faliciparum* out of the population under NMEP's surveillance. Malaria data is only available from 1975 and 1995.

Rainfall data comes from the University of Delaware's Center for Climatic Research, which provides rainfall estimates at 0.5° longitude and latitude nodes across the world. A rainfall shock is measured as the fractional deviation of rainfall from its average level (calculated over 1916–1995), summed over all months. We also add temperature data to this dataset. Data are taken from the (India Water Portal 2020). This data is derived from the publicly available Climate Research Unit TS2.1 dataset from the Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia in Norwich, UK. Temperature shocks are defined using the same methodology described above for rainfall shocks. Data on district area, river kilometers, elevation, overall gradient, and river gradients come from CIESIN, Earth Institute Columbia University.

As in previous studies relying on India data (e.g., Sarsons 2015; Duflo and Pande 2007b), the unit of analysis is the district as the most disaggregated data is only available at this level. A district is an administrative unit immediately under a state (Census 1981 used as a reference). Duflo and Pande (2007b) argue that the district in which a dam is located includes most or all of its catchment area, and a portion of the command area, while most of the command area is in the downstream district. They suggest that the effect of a dam will be unambiguously positive downstream (and likely underestimated). By contrast, it is not clear what to expect overall for the district in which a dam is located (as the positive 'command' effects could dominate the negative 'catchment' effects, or vice versa).

4. Empirical Results

4.1. The Importance of Geography in Large Dam Construction

Table 1 displays the G2SIV estimation results of Equation (2) in column 1, where the number of dams in the district is the dependent variable and regressed against river gradients and a set of geographical variables that are all interacted with the predicted dam incidence at the state level. District fixed effects, state-year fixed effects, and common factors are included. The estimation includes 358 districts between 1975 and 1995.

Table 1. The importance of geography in large dam construction.

	Interacted with Predicted Number of Dams in the State
Number of dams	(1)
Fraction river gradient 1.5–3%	11.619 ***
	(4.117)
Fraction river gradient 3–6%	−12.671 *
	(7.622)
Fraction river gradient above 6%	11.785 ***
	(2.436)
River length	−0.000
	(0.001)
District area	0.000 ***
	(0.000)
Elevation 250–500 m	−0.414
	(1.025)
Elevation 500–1000 m	2.881 *
	(1.484)
Elevation above 1000 m	−4.771
	(5.661)
District gradient 1.5–3%	10.500 *
	(6.073)
District gradient 3–6%	3.528
	(9.504)
District gradient above 6%	1.126
	(5.792)
F-test for river gradient, *p*-val. (stat.)	0.000 (97.38)
N	6937
Number of districts	358
Common factors	YES
District FE	YES
Rho	0.763
J-test	0.53
Number of instruments	44
Number of common factors	1
Number of lags	3

Notes: Robust standard errors in parentheses. *** *p*-val. <0.01, ** *p*-val. <0.05, * *p*-val. <0.1, # *p*-val. < 0.15. All regressors are interacted with predicted dam incidence at the state level. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a full sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 358 districts.

All covariates are defactored based on a common set of factors estimated jointly. We use 3 lags of all covariates as defactored instruments, up to a maximum of 1 factor following the Ahn and Horenstein (2013) criteria. We control for district-specific and state-year fixed effects by eliminating them prior to the estimation. The *p*-value of the J test statistic (0.53) indicates that the over-identifying restrictions (instruments) are valid. The estimated number of factors in the first and second steps both equal 1.

The rho is about 0.76 in column 1, suggesting that most of the variation in the composite error term is due to the single unobserved factor, conditional on district fixed effects, and state-year fixed effects. Put simply, estimators that do not control for common shocks are likely to be severely biased. In essence, our results in column 1 mimic those of Duflo and Pande (2007b) and support the importance of engineering considerations for dam construction. We find that a gentle river gradient (1.5–3%) and a very steep river gradient (3–6%) increase the number of dams. Interestingly, a steep river gradient (3–6%) has a negative effect on dam construction. The *p*-value for the F-test for the joint significance of all river gradient variables is 0.00 (F-stat: 97.38). In addition, the district area seems to be a positive factor in building dams in upstream districts.

4.2. The Impact of Large Dams on Malaria Incidence

Table 2 displays G2SIV estimates of the impact of large dams on malaria incidence both in the district where they are located ("Dams") and downstream ("Upstream dams") for the baseline model. Columns 1, 2, and 3, respectively, display G2SIV estimates where the number of common factors changes from 3 to 1. We rely on several selection criteria to choose the number of common factors and they suggest this number to be either 1 (based on Ahn and Horenstein 2013) or 3 (based on Onatski 2010).

Table 2. Baseline results.

Malaria, log	**(1)**	**(2)**	**(3)**
Malaria, log, $t-1$	0.654 ***	0.523 ***	0.524 ***
	(0.006)	(0.007)	(0.009)
Dams	0.913 ***	1.096 ***	1.461 ***
	(0.301)	(0.349)	(0.401)
Upstream dams	1.085 ***	1.462 ***	1.041 **
	(0.309)	(0.384)	(0.430)
N	2845	2845	2845
Number of districts	355	355	355
Geographical controls	YES	YES	YES
Common factors	YES	YES	YES
District FE	YES	YES	YES
Rho	0.396	0.355	0.213
J test	0.582	0.429	0.356
Number of instruments	140	140	140
Number of common factors	3	2	1
Number of lags	6	6	6

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables.

In our baseline specification, the two dam variables are defactored based on a common set of factors estimated separately, based on its own estimated factors, from the one for all the other variables. We use 6 lags of covariates as defactored instruments, and up to a maximum of 3 factors. The number of lags is chosen to minimize the variance of the model's standard error. We control for district-specific and state-year fixed effects by eliminating them prior to the estimation. In our baseline model, the estimated number of factors in the first step equals 1 throughout the columns. In robustness analyses, we test whether our main results hold to alternative specifications. We discuss these results later.

Finally, as explained earlier, all columns use the predicted number of dams (in their own district and downstream) as instruments for the actual number of dams. The J test statistic in column 1 does not reject the null hypothesis as the p-value equals 0.582. This implies that the additional instruments, i.e., predicted numbers of dams in own district and upstream, may be valid instruments.

The rho across columns in Table 2 is between 0.2–0.4, suggesting that a substantial part of the variation in the composite error term is due to the single unobserved factor, conditional on district fixed effects and state-year fixed effects. Put simply, estimators that do not control for common shocks may be likely biased.

The main message from Table 2 is that an increase in the number of dams in a district increases malaria incidence both in the district and in downstream areas. Our baseline estimate in column 1 implies that a one-dam increase in a district would increase malaria incidence by about 0.9 percent. Overall, according to columns (1)–(3), a one-dam increase in a district would increase malaria incidence by between 0.9 to 1.4 percent. The construction of a dam in the district likely creates additional breeding grounds for Anopheles mosquitoes that facilitate the development of malaria transmission.

Furthermore, we find some evidence that dams also increase malaria incidence in downstream districts in all columns. According to columns (1)–(3), a one-dam increase in a district would increase malaria incidence by between 1 to 1.5 percent. We assume this could be due to agricultural development in downstream areas. We will examine potential agricultural mechanisms later in the paper.

These findings are the first evidence of the link between dams and malaria incidence in the empirical literature (that accounts for the endogeneity of dam location). While our results are in contrast with Duflo and Pande (2007b) who find no effect of dams on malaria incidence, they fit the increasing body of anecdotal evidence suggesting dams have been associated with malaria (e.g., Kibret et al. 2019b). Last, the coefficients for the lagged dependent variable are statistically significant and positive in all columns, and range from 0.5 to 0.6, suggesting some persistence of malaria.

Next, we examine how our findings change if we change the number of lags in the G2SIV estimation. Table 3 confirms our baseline findings on the impact of dam construction on malaria in the district in which a dam is located. According to columns (1)–(6), a one-dam increase in a district would increase malaria incidence by approximately 0.9 to 1.5 percent. The impact of dam construction on malaria in downstream areas seems to be more sensitive to the number of lags used in the IV estimation. Columns 1 to 4 suggest there is no impact of a dam on malaria in downstream areas. If we use further lags in columns 5 and 6, the estimates confirm our baseline results. As explained earlier, the lag selection is chosen to minimize the variance of the standard error, so our baseline results are our preferred specification.

Table 3. Changing the number of lags used in the G2SIV estimation.

Malaria, log	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Malaria, log, $t-1$	0.713 ***	0.702 ***	0.606 ***	0.605 ***	0.536 ***	0.539 ***
	(0.021)	(0.022)	(0.013)	(0.013)	(0.2165)	(0.001)
Dams	0.883 ***	1.374 ***	0.892 **	0.945 **	1.284 ***	1.541 ***
	(0.293)	(0.456)	(0.436)	(0.444)	(0.289)	(0.215)
Upstream dams	0.195	−0.658 #	0.443	0.332	1.135 ***	1.620 ***
	(0.416)	(0.420)	(0.406)	(0.375)	(0.263)	(0.237)
N	4177	4177	3508	3508	2186	2186
Number of districts	357	357	356	356	352	352
Geographical controls	YES	YES	YES	YES	YES	YES
Common factors	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES
Rho	0.142	0.350	0.166	0.396	0.221	0.467
J test	0.218	0.386	0.046	0.040	0.435	0.530
Number of instruments	100	100	120	120	160	160
Number of common factors	1	3	1	3	1	3
Number of lags	4	4	5	5	7	7

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables.

4.3. Alternative Specifications

We test the robustness of our baseline findings to several alternative specifications. Table A2 in Appendix A provides estimates from extended and reduced-form models. In the extended model, we include rainfall and temperature shocks, as is sometimes done in the literature. The use of reduced-form models allows examining whether bad controls drive our estimates. Across columns, the "dams" estimates are positive and statistically significant (except for the dam estimate in column 4 which is somewhat imprecise). The range of estimates implies a one-dam increase would increase malaria prevalence by about

0.5–1.5 percent. Furthermore, when we look at the downstream effect of dams on malaria, we find evidence in support of our baseline results. In columns 1 and 2 (extended models), we find that dams do have a downstream effect on malaria. Similarly, the "upstream dams" estimates are positive, and statistically significant, and are respectively 1.145 and 1.192 in columns 5 and 6.

Table A3 drops years in the estimation sample, as common factor approaches can be somewhat sensitive to sample selection. Columns 1 and 2 (3 and 4) exclude the first 3 (last 3) years of the dataset. In most cases, the estimates are in line with that of Table 2. They imply a one-dam increase in the number of dams in a district would increase malaria prevalence by about 0.9–1.9 percent. We find less consistent evidence for the downstream effect of dams on malaria. For instance, the "upstream dam" estimate in column 2 is however negative and statistically insignificant.

Table A4 uses one to five-year lagged dams and upstream dams variables in the baseline model, instead of contemporaneous dam variables. This specification examines whether there are large lags in the effect of dam construction on malaria, especially in downstream areas. There are two reasons why this specification is necessary. First, there might have been discrepancies in the administrative reporting with respect to the end of dam construction. Perhaps, local administrators have an incentive to report an earlier end date than the actual end date to appear compliant with expectations from federal or state authorities that often mainly contribute to the funding of dam infrastructures. Second, it may be important to consider that the spatial disease spread is likely going downstream due to the flight paths of Anopheles mosquitoes, and thus the lagged specification can inform how dam construction can have delayed effects on malaria transmission, especially in downstream areas, as we consider the likely disease spread across the geography of districts over time.

The estimates from Table A4 are particularly interesting because they confirm the malaria-increasing effect of dams persists over time in the district where the dam is located, as well as in downstream areas. In Table A5, we present estimation results relying on a slightly modified model than the one presented in Table 1. In particular, the number of common factors is equal to 4, following the selection criteria from Onatski (2010). Table A5 shows the main pattern of results remains the same. The dams coefficients remain positive, statistically significant, and in line with estimates from Table 2.

4.4. Agricultural Mechanisms

What is the contribution of agricultural development to the increase in malaria incidence due to dam construction? Table 4 examines several pathways between the dam and agricultural development. In particular, Table 4 estimates the impact of dam construction on several agricultural outcomes, i.e., agricultural production (i.e., value of the production of main crops, i.e., rice, wheat, sugarcane, jowar, bajra, maize), yield of the main crops, gross irrigated area, gross cultivated area, and high yielding variety (HYV) crop cultivated area. Sample size varies across these outcomes but typically starts earlier than 1975. More details can be found in the notes below in Table 4.

First, we find that dam construction benefits agriculture in the areas in which they are built. For example, the "dams" estimates in columns 1 and 2 are positive and statistically significant. This suggests that a one-dam increase in a district would increase agricultural production by about 0.2 percent.

The increase in agricultural production seems to be driven by the dam construction effect on irrigated and cultivated areas. A one-dam increase in a district would increase irrigated (cultivated) areas by about 0.2–0.3 (0.1) percent. Put simply, dams expand the role of irrigation as an agricultural practice in nearby farmlands in our context. The extent to which irrigation systems lack efficient drainage amplifies the existence of additional breeding grounds for Anopheles mosquitoes.

Table 4. Agricultural mechanisms.

Dependent Variable	**Agricultural Production, log**		**Yield, log**	**Irrigated Area, log**		**Cultivated Area, log**	**HYV Area, log**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Lagged dependent variable	0.435 ***	0.428 ***	0.105 ***	0.947 ***	0.547 ***	0.522 ***	0.381 ***	0.396 ***
	(0.011)	(0.010)	(0.0066)	(0.037)	(0.001)	(0.010)	(0.010)	(0.008)
Dams	0.232 **	0.247 **	0.198 #	0.207 **	0.314 ***	0.107 ***	0.022	0.036
	(0.090)	(0.096)	(0.136)	(0.0834)	(0.051)	(0.030)	(0.090)	(0.082)
Upstream dams	0.286 ***	0.361 ***	0.258 **	−0.003	−0.045	−0.098 ***	0.566 ***	0.597 ***
	(0.072)	(0.068)	(0.102)	(0.076)	(0.054)	(0.030)	(0.106)	(0.101)
N	5561	5561	5561	5280	3912	3902	3882	3882
Number of districts	266	266	266	266	266	265	264	264
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES
Common factors	YES	YES	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES	YES	YES
Rho	0.224	0.355	0.377	0.211	0.639	0.642	0.292	0.516
p-val J test	0.176	0.642	0.075	0.440	0.403	0.298	0.725	0.578
Number of instruments	120	120	120	40	140	140	140	140
Number of common factors	1	3	1	1	2	1	1	3
Number of lags	5	5	5	1	6	6	6	6

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. The sample includes 268 districts, instead of 358, due to missing data. Note these estimations are based on a sample ranging from 1966 to 1995, but the sample size varies across columns due to missing data. All G2SIV estimates. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables. Lag selection is determined to minimize the variance of the model's standard error. Description of agricultural outcomes used in Table 4 (source: Duflo and Pande 2007b). Agricultural production: the value of the production of main crops, includes rice wheat, sugarcane, jowar (sorghum), bajra (pearl millet), maize in rupees per thousand tons (average prices for crops 1960–65 are used). Yield: yield of the main crops: rice wheat, sugarcane, jowar (sorghum), bajra (pearl millet), maize, the yield being measured as the number of rupees per hectare (average prices for crops 1960–65 are used). Gross irrigated area: gross irrigation area, in thousand hectares. Gross cultivated area: gross cultivation area being measured in thousands hectares. HYV area: area under HYV crop cultivation.

These additional results portray a very different picture than the previous literature that often finds only negative or no effect on agriculture in the vicinity of dams (e.g., Strobl and Strobl 2011; Blanc and Strobl 2014; for a review, see Dillon and Fishman 2019). In particular, our results directly contradict Duflo and Pande (2007b) in the India context, but are also in line with more descriptive analyses (e.g., Singh 1990; Pradhan and Srinivasan 2022). To contextualize our results in the face of the literature, let's note that all the above-mentioned studies typically rely on state-year fixed effects in their modeling. We argued earlier why this modeling choice may lead to biased estimations.

In addition, in downstream areas, dam construction increases agricultural production (0.2–0.4%), yield (0.3%), and the cultivation of HYV crops (0.6%). The existence of positive benefits in downstream areas is in line with the literature. However, we do not find that irrigated areas or cultivated area[11] respond to dam construction, so the benefits seem to be driven by changes in cropping patterns. In particular, the cultivation of HYV crops increases. These cropping patterns can arguably be conducive to the reproduction of mosquitoes and may explain the increase in malaria incidence in downstream areas.

4.5. A Back-of-the Envelope Comparison of the Malaria Costs and Agricultural Production Gains

Finally, we present an additional set of empirical results that might shed light on quantifying the costs of malaria and the agricultural production gains resulting from dam construction, because malaria will lead to illnesses and deaths that eventually reduce productivity and production. In essence, we use the same model as displayed in Table 4 column 1 but we now include malaria incidence in the model as an independent variable. We account for the endogeneity of malaria by using 3 lags of malaria incidence as instruments. Results can be found in Table A6 in Appendix A. Following column 1 (preferred specification), we find that a 1% increase in malaria reduces agricultural production by 0.007%. Note the coefficient is significant at 10.9 percent, so marginally insignificant at 10 percent. We also find that a one-dam increase in a district (upstream) would increase agricultural production by about 0.3 (0.5) percent, in line with Table 4. It is noteworthy that the malaria coefficient is statistically different from the dams and upstream dams coefficients.

If we combine these results with our baseline estimates in Table 2 (that is, a one-dam construction in a district would result in about 0.9–1.5 percent increase in malaria), the estimate from Table A6 column 1 would suggest that the agricultural gains dominate the losses from malaria (expressed in agricultural production).

However, it is important to remind the reader that the estimates from Tables 2 and A6 are based on different samples, and therefore any comparison should be taken with caution. Similarly, we do not take into account the fact that gains and losses may differ over time; this would arguably affect any comparison of gains and losses.

5. Conclusions

The empirical literature on the links between large irrigation dams and malaria incidence mainly consists of case studies and anecdotal evidence. The few econometric studies regarding this topic remain vulnerable to several biases, among which the endogeneity bias due to the non-random nature of dam allocation and the omitted variable bias due to (unaccounted) time-varying heterogeneity.

Once we account for these biases, we find that large dams increase malaria incidence in districts where they are located, as well as in downstream areas. Our estimates imply a one-dam increase in the number of dams in the district increases annual malaria incidence by about 0.9 to 1.4 percent, while the effect on malaria is about 1 to 1.5 percent in downstream districts. This paper provides evidence that large irrigation dams increase malaria incidence and confirms an increasing literature based on case studies and anecdotal reports. We also find this malaria-increasing effect of dams lasts over time.

Furthermore, we examine the effect of dam construction on agricultural outcomes to highlight potential mechanisms between dam and malaria incidence. We find that dam construction benefits agriculture in the vicinity of dams, as well as in downstream areas. These positive effects are driven by increased irrigation in the vicinity of dams, while they are driven by changes in cropping patterns in downstream areas, where the cultivation of HYV crops (with larger water requirements) increases.

Both results related to malaria incidence and agricultural outcomes are novel in the empirical literature and should substantially alter future distributional cost-benefit analyses on large dam construction. They have clear implications for policymakers and researchers, especially within the context of the SDG agenda. From a policy standpoint, the construction of dams in malaria-sensitive regions should be coupled with targeted interventions, such as the wide deployment of insecticide-treated nets or the roll-out of future vaccines. From a methodological perspective, our paper highlights the role of time-varying heterogeneity in empirical studies analyzing the impact of irrigation infrastructures. Our results suggest the use of common factors approaches can be beneficial to researchers.

Finally, it is important to state one limitation of our paper. Our results are based on a period 1975–1995 in which the government of India was implementing a specific set of actions and interventions to control and eradicate malaria. It is noteworthy that government actions to reduce and eradicate malaria have substantially changed over time in India. The period covered by the dataset starts at a time when India was facing a resurgence of malaria. After the launching of the National Malaria Control Programme (NMCP) in 1953 and the National Malaria Eradication Programme (NMEP) in 1958, substantial progress had been reached in the 1950–1960s, and deaths due to malaria were completely eliminated (Sharma 1998).

However, in the early 1970s, a sense of complacency following the successful efforts of the 1950s–1960s, dichlorodiphenyltrichloroethane (DDT) shortages, and the increasingly widespread mosquito's resistance to DDT used for spraying and the parasite's resistance to anti-malaria drugs led to a resurgence of malaria cases and deaths countrywide. The national programmes continued to face technical, financial, and administrative problems during the period covered by the dataset, including insecticide shortages, refusal for spraying, insecticide resistance, inadequate surveillance, and dismantling of research on malaria (Sharma 1998).

After the mid-90s, authorities shifted their focus from the eradication and control of malaria to the prevention, early detection, and treatment of human cases. They have progressively relied on more targeted indoor spraying, biological control via the use of larvivorous fish and bio-larvicides, and personal protection strategies, mainly insecticide treated bed nets. Government actions have also been increasingly targeting vulnerable populations and malaria-endemic areas. India's efforts have resulted in a substantial decline in cases and deaths over the last two decades.

Given this discussion, it is likely that the link between dam construction and malaria is now weaker than it was over the period covered by the dataset. However, India's improvement in the face of the disease has required large public investments and it is not clear whether other countries, especially in Africa, have the public finances to support such policies.

Author Contributions: Conceptualization, S.M.; Methodology, S.M.; Software, S.M., K.C. and S.S.; Formal analysis, S.M. and A.S.; Writing, S.M., K.C. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in the dataverse referenced in the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Descriptive statistics of baseline sample.

	Obs.	Mean	Std. Dev.	Min.	Max.
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Malaria incidence	2845	3.725	6.711	0	95.51
Malaria incidence, log	2845	0.0753	1.814	−4.605	4.559
Rainfall shock	2012	0.0461	0.241	−0.773	1.281
Temperature shock	1911	0.00473	0.311	−1.127	1.259
Number of dams in district	2845	7.682	14.24	0	90
Number of dams upstream	2845	13.53	27.99	0	246.0
Elevation 250–500 m	2845	0.846	1.636	0	7.348
Elevation 500–1000 m	2845	0.513	1.228	0	7.680
Elevation above 1000 m	2845	0.0483	0.263	0	2.806
Fraction of river gradient 1.5–3%	2845	0.193	0.301	0	1.893
Fraction of river gradient 3–6%	2845	0.118	0.240	0	2.489
Fraction of river gradient above 6%	2845	0.127	0.356	0	2.720
District area	2845	22,100	30,552	0	288,978
River length	2845	1503	3720	0	46,303
Fraction of district gradient 1.5–3%	2845	0.273	0.426	0	2.414
Fraction of district gradient 3–6%	2845	0.160	0.289	0	2.159
Fraction of district gradient above 6%	2845	0.147	0.353	0	2.746

Table A2. Robustness analyses: alternative models.

	Extended Model		Reduced form Models			
Malaria, log	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Malaria, log, t − 1	0.541 ***	0.605 ***	0.542 ***	0.677 ***	0.508 ***	0.635 ***
	(0.003)	(0.002)	(0.010)	(0.718)	(0.009)	(0.006)
Rainfall shock	0.068 ***	0.039 *				
	(0.022)	(0.020)				
Temperature shock	0.113 ***	0.069 ***				
	(0.010)	(0.017)				
Dams, own district	0.586 **	0.865 ***	1.543 ***	0.486 #		
	(0.256)	(0.255)	(0.434)	(0.326)		
Upstream dams	0.650 ***	0.821 ***			1.145 **	1.192 ***
	(0.246)	(0.185)			(0.447)	(0.312)
N	1886	1886	2845	2845	2845	2845
Number of districts	241	241	355	355	355	355
Geographical controls	YES	YES	YES	YES	YES	YES
Common factors	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES
Rho	0.192	0.430	0.202	0.388	0.221	0.404
J test	0.520	0.322	0.397	0.673	0.570	0.562
Number of instruments	154	154	126	126	126	126
Number of factors	1	3	1	3	1	3
Number of lags	6	6	6	6	6	6

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables.

Table A3. Robustness analyses: dropping years in the estimation sample.

	1978–1995	**1978–1995**	**1975–1992**	**1975–1992**
Malaria, log	**(1)**	**(2)**	**(3)**	**(4)**
Malaria, log, t − 1	0.422 ***	0.441 ***	0.524 ***	0.654 ***
	(0.002)	(0.000)	(0.009)	(0.006)
Dams	1.884 ***	0.992 ***	1.461 ***	0.913 ***
	(0.534)	(0.227)	(0.401)	(0.301)
Upstream dams	0.415 *	−0.056	1.041 **	1.085 ***
	(0.2176)	(0.090)	(0.430)	(0.309)
N	1907	1907	2845	2845
Number of districts	354	354	355	355
Geographical controls	YES	YES	YES	YES
Common factors	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Rho	0.290	0.639	0.213	0.369
J test, p-val.	0.259	0.622	0.356	0.582
Number of instruments	140	140	140	140
Number of common factors	1	3	1	3
Number of lags	6	6	6	6

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables. Columns 1 and 2 drop the first 3 years of the sample, while columns 3 and 4 drop the last 3 years of the sample.

Table A4. Robustness analyses: Lagging the dams coefficients.

	t − 1		t − 2		t − 3		t − 4		t − 5	
Malaria, log	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Malaria, log	0.699 ***	0.562 ***	0.717 ***	0.571 ***	0.531 ***	0.484 ***	0.553 ***	0.579 ***	0.676 ***	0.670 ***
	(0.009)	(0.014)	(0.015)	(0.018)	(0.019)	(0.032)	(0.020)	(0.044)	(0.074)	(0.089)
Dams, lagged	0.589 #	1.779 ***	0.746 #	1.193 #	2.264 ***	0.030 ***	1.693 ***	1.342 *	0.121	0.894 *
	(0.401)	(0.458)	(0.499)	(0.7581)	(0.419)	(0.009)	(0.398)	(0.725)	(0.419)	(0.459)
Upstream dams, lagged	0.715 **	0.836 *	0.978 **	0.823 *	1.294 ***	0.021 ***	0.795 #	1.612 **	0.980 **	1.413 **
	(0.285)	(0.446)	(0.380)	(0.499)	(0.446)	(0.006)	(0.494)	(0.663)	(0.406)	(0.628)
N	2845	2845	2845	2845	2483	3144	2141	2798	3115	3778
Number of districts	355	355	355	355	353	356	352	354	355	356
Geographical controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Common factors	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Rho	0.387	0.201	0.383	0.200	0.447	0.231	0.514	0.169	0.373	0.148
p-val J test	0.632	0.560	0.559	0.266	0.418	0.254	0.663	0.319	0.479	0.405
Number of instruments	120	120	100	100	100	80	100	80	60	40
Number of common factors	3	1	3	1	3	1	3	1	3	1
Number of lags	5	5	4	4	4	3	4	3	2	1

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables. Dams coefficients are replaced by their lagged values. For example, columns 1 and 2 lag the dams coefficients by one period, while column 9 and 10 lag the dams coefficients by 5 periods.

Table A5. Robustness analyses: baseline results based on different estimation than in Table 1 (change in common factors to 4).

Malaria, log	**(1)**	**(2)**
Malaria, log, t − 1	0.646 ***	0.514 ***
	(0.006)	(0.009)
Dams	0.853 ***	1.350 ***
	(0.307)	(0.401)
Upstream dams	1.106 ***	0.946 **
	(0.309)	(0.420)
N	2845	2845
Number of districts	355	355
Geographical controls	YES	YES
Common factors	YES	YES
District FE	YES	YES
Rho	0.398	0.385
J test	0.555	0.218
Number of instruments	140	140
Number of common factors	3	1
Number of lags	6	6

Notes: Robust standard errors in parentheses. *** p-val. < 0.01, ** p-val. < 0.05, * p-val. < 0.1, # p-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. The sample includes 355 districts, instead of 358, due to missing data. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables.

Table A6. Evaluating the costs of malaria on agricultural production.

Dependent Variable	**Agricultural Production, log**				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Lagged dependent variable,	0.241 ***	0.107 ***	0.171 ***	0.212 ***	0.272 ***
	(0.069)	(0.037)	(0.024)	(0.018)	(0.013)
Dams	0.302 *	0.349 **	0.276 *	0.220 *	0.253 **
	(0.177)	(0.170)	(0.141)	(0.120)	(0.105)
Upstream dams	0.481 ***	0.705 ***	0.647 ***	0.629 ***	0.532 ***
	(0.153)	(0.145)	(0.001)	(0.112)	(0.090)
Malaria, log	−0.007 #	−0.002	−0.001	−0.006 *	−0.004 #
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
N	3025	2943	2897	2894	2891
Number of districts	261	261	261	261	261
Geographical controls	YES	YES	YES	YES	YES
Common factors	YES	YES	YES	YES	YES
District FE	YES	YES	YES	YES	YES
Rho	0.398	0.401	0.395	0.390	0.391
p-value, J test	0.418	0.093	0.122	0.113	0.091
Number of instruments	64	84	104	124	144
Number of common factors	1	1	1	1	1
Number of lags	2	3	4	5	6

Notes: Robust standard errors in parentheses. *** *p*-val. < 0.01, ** *p*-val. < 0.05, * *p*-val. < 0.1, # *p*-val. < 0.15. Dams and upstream dams coefficients are multiplied by 100. Rho is the fraction of the variance of the error term that is explained by the factor component. The sample includes 261 districts, instead of 358, due to missing data. Note these estimations are based on a sample ranging from 1975 to 1995. All G2SIV estimates. All geographical variables (interacted with predicted dam incidence at the state level) from Table 1 are included except for the river gradient variables. Lag selection is determined to minimize the variance of the model's standard error.

Notes

1. While the largest share of the malaria burden falls on Africa, the majority of people living in close proximity of the reservoirs of large dams in malaria-endemic areas are in India.
2. Koka dam, Kesem dam, Koga dam.
3. However, they also report that the population of adult Anopheles mosquitoes increased in villages closer to a dam than those further away.
4. The non-random location of dams may confound estimates. For example, local governments may target areas whose agriculture is either already productive.
5. The case for reliable research on large dams is even more compelling because of their controversial nature (e.g., Duflo and Pande 2007b; Strobl and Strobl 2011; Dillon and Fishman 2019).
6. Crops that have been bred or fertilized and can be produced by genetic modifications to increase the rate of production. They are more resistant to insects and disease and have played a key role in India's Green Revolution.
7. It could also be lower as wealthier districts may be better prepared to fight the disease with larger human, financial, and medical, infrastructures and resources.
8. This strategy (or a variant of) has been used implicitly or explicitly in Strobl and Strobl (2011), Sarsons (2015), Blanc and Strobl (2014) or Mettetal (2019).
9. To validate the use of factor models, we use the cross section dependence test from Pesaran (2015). The hypothesis of null cross section dependence is rejected at the 1 per cent level in our application. This suggests that the presence of common factors influencing malaria across districts and explicitly validates the factor model approach.
10. We use the Stata command xtivdfreg designed by Kripfganz and Sarafidis (2021) to implement the approach, and the command xtnumfac from Ditzen and Reese (2022).
11. The negative effect on cultivated area is economically small.

References

Ahn, Seung C., and Alex R. Horenstein. 2013. Eigenvalue Ratio Test for the Number of Factors. *Econometrica* 81: 1203–27.

Asenso-Okyere, Kwadwo, Felix A. Asante, Jifar Tarekegn, and Kwaw S. Andam. 2009. *The Linkages between Agriculture and Malaria: Issues for Policy, Research, and Capacity Strengthening*. IFPRI Discussion Paper. Washington: IFPRI.

Asfaw, Solomon, Alessandro Carraro, Benjamin Davis, Sudhanshu Handa, and David Seidenfeld. 2017. Cash transfer programmes, weather shocks and household welfare: Evicence from a randomized experiment in Zambia. *Journal of Development Effectiveness* 9: 419–42. [CrossRef]

Bai, Jushan. 2009. Panel data models with interacted fixed effects. *Econometrica* 77: 1229–79.

Bai, Jushan, and Kunpeng Li. 2014. Theory and methods of panel data models with interactive effects. *Annals of Statistics* 42: 142–70. [CrossRef]

Blanc, Elodie, and Eric Strobl. 2014. Is small better? A comparison of the effect of large versus small dams on cropland productivity in South Africa. *The World Bank Economic Review* 28: 545–76. [CrossRef]

Cui, Guowei, Milda Norkutė, Vasilis Sarafidis, and Takashi Yamagata. 2022. Two-stage instrumental variable estimation of linear panel data models with interactive effects. *The Econometrics Journal* 25: 340–61. [CrossRef]

Dejenie, Tadesse, Mekonnen Yohannes, and Tsehaye Assmelash. 2012. Adult mosquito populations and their health impact around and far from dams in Tigray Region, Ethiopia. *Momona Ethiopia Journal of Science* 4: 40–51.

Dillon, Andrew, and Ram Fishman. 2019. Dams: Effects of Hydrological Infrastructure on Development. *Annual Review of Resource Economics* 11: 125–48. [CrossRef]

Ditzen, Jan, and S. Reese. 2022. *xtnumfac*: A battery of estimators for the number of common factors in time series and panel data models. Paper presented at 28th UK Stata Conference 2022, London, UK, September 8–9.

Duflo, Esther, and Rohini Pande. 2007a. Dams, Poverty, Public Goods and Malaria Incidence in India. Available online: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MNIBOL (accessed on 1 September 2019). [CrossRef]

Duflo, Esther, and Rohini Pande. 2007b. Dams. *The Quarterly Journal of Econometrics* 122: 601–46. [CrossRef]

Falavigna-Guilherme, Ana Lucia, Allan Martins da Silva, Edson Valdemar Guilherme, and Dina Lúcia Morais. 2005. Retrospective study of malaria prevalence and Anopheles genus in the area of influence of the binational Itaipu Reservoir. *Journal of the São Paulo Institute of Tropical Medicine* 47: 81–86. [CrossRef]

FitzHugh, Thomas W., and Richard M. Vogel. 2011. The impact of dams on flood flows in the US. *River Research and Applications* 27: 1192–215. [CrossRef]

GBCHealth. 2012. *Linkages between Malaria and Agriculture*. New York: GBCHealth.

Ghebreyesus, Tedros A., Mitiku Haile, Karen H. Witten, Asefaw Getachew, Ambachew M. Yohannes, Mekonnen Yohannes, Hailay D. Teklehaimanot, Steven W. Lindsay, and Peter Byass. 1999. Incidence of malaria among children living near dams in northern Ethiopia: Community-based incidence survey. *The British Medical Journal (Clinical Research Ed.)* 3197211: 663–66. [CrossRef] [PubMed]

Gobillon, Laurent, and Thierry Magnac. 2016. Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls. *The Review of Economics and Statistics* 98: 535–51. [CrossRef]

Guerra, C. A., R. W. Snow, and S. I. Hay. 2006. A global assessment of closed forests, deforestation and malaria risk. *Annals of Tropical Medicine and Parasitology* 100: 189–204. [CrossRef]

India Water Portal. 2020. Available online: https://www.indiawaterportal.org/ (accessed on 1 January 2020).

International Water Management Institute. 2018. *Dams and Malaria in Africa: Time for Action*. (IWMI Water Policy Brief 40). Colombo: International Water Management Institute (IWMI). 8p.

Janko, Mark M., Seth R. Irish, Brian J. Reich, Marc Peterson, Stephanie M. Doctor, Melchior Kashamuka Mwandagalirwa, Joris L. Likwela, Antoinette K. Tshefu, Steven R. Meshnick, and Michael E. Emch. 2018. The links between agriculture, Anopheles mosquitoes, and malaria risk in children younger than 5 years in the Democratic Republic of the Congo: A population-based, cross-sectional, spatial study. *Lancet Planetary Health* 2: 74–82. [CrossRef]

Keiser, Jennifer, M. Caldas De Castro, Michael F. Maltese, Robert Bos, Marcel Tanner, Burton H. Singer, and Jurg Utzinger. 2005. Effect of irrigation and large dams on the burden of malaria on a global and regional scale. *The American Journal of Tropical Medicine and Hygiene* 72: 392–406. [CrossRef]

Kibret, Solomon. 2018. Time to revisit how dams are affecting malaria transmission. *Lancet Planetary Health* 2: 378–79. [CrossRef]

Kibret, Solomon, G. Glenn Wilson, Darren Ryder, Habte Tekie, and Beyene Petros. 2019a. Environmental and meteorological factors linked to malaria transmission around large dams at three ecological settings in Ethiopia. *Malaria Journal* 18: 54. [CrossRef]

Kibret, Solomon, Jonathan Lautze, Matthew McCartney, Luxon Nhamo, and Guiyun Yan. 2019b. Malaria around large dams in Africa: Effect of environmental and transmission endemicity factors. *Malaria Journal* 18: 303. [CrossRef]

Kripfganz, Sebastian, and Vasilis Sarafidis. 2021. Instrumental-variable estimation of large-T panel-data models with common factors. *The Stata Journal* 21: 659–86. [CrossRef]

Kuriakose, Anne T., Rasmus Heltberg, William Wiseman, Cecilia Costella, Rachel Cipryk, and Sabine Cornelius. 2013. Climate-Responsive Social Protection. *Development Policy Review* 31: 19–34. [CrossRef]

Larinier, Michel. 2001. *Dams, Fish and Fisheries: Opportunities, Challenges and Conflict Resolution*. Rome: The Food and Agriculture Organization.

Lautze, Jonathan, Matthew McCartney, Paul Kirshen, Dereje Olana, Gayathree Jayasinghe, and Andrew Spielman. 2007. Effect of a large dam on malaria risk: The Koka reservoir in Ethiopia. *Tropical Medicine and International Health* 12: 982–89. [CrossRef] [PubMed]

Lee, Gwanjae, Hye Won Lee, Yong Seok Lee, Jung Hyun Choi, Jae E. Yang, Kyoung Jae Lim, and Jonggun Kim. 2019. The Effect of Reduced Flow on Downstream Water Systems Due to the Kumgangsan Dam under Dry Conditions. *Water* 11: 739. [CrossRef]

Lerer, Leonard B., and Thayer Scudder. 1999. Health impacts of large dams. *Environmental Impact Assessment Review* 19: 113–23. [CrossRef]

MacDonald, Andrew J., and Erin A. Mordecai. 2010. Amazon deforestation drives malaria transmission, and malaria burden reduces forest cleaning. *Proceedings of the National Academy of Sciences* 116: 22212–18. [CrossRef]

Mary, Sebastien. 2022. Dams mitigate the effect of rainfall shocks on Hindus-Muslims riots. *World Development* 150: 105731. [CrossRef]

Mbakop, Lili R., Parfait H. Awono-Ambene, Stanislas E. Mandeng, Wolfgang E. Ekoko, Betrand N. Fesuh, Christophe Antonio-Nkondjio, Jean-Claude Toto, Philippe Nwane, Abraham Fomena, and Josiane Etang. 2019. Malaria Transmission around the Memve'ele Hydroelectric Dam in South Cameroon: A Combined Retrospective and Prospective Study, 2000–16. *International Journal of Environmental Research and Public Health* 16: 1618. [CrossRef]

McCartney, Matthew P., and Hilmy Sally. 2007. Managing the Environmental Impacts of Dams. Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ee489330c82d80028a08c7856b1d9a6d2d99576d (accessed on 18 June 2023).

Mettetal, Elizabeth. 2019. Irrigation dams, water and infant mortality: Evidence from South Africa. *Journal of Development Economics* 138: 17–40. [CrossRef]

Ndiath, Mamadou O., Jean-Biram Sarr, Lobna Gaayeb, Catherine Mazenot, Seynabou Sougoufara, Lassana Konate, Franck Remoue, Emmanuel Hermann, Jean-francois Trape, Gilles Riveau, and et al. 2012. Low and seasonal malaria transmission in the middle Senegal River basin: Identification and characteristics of Anopheles vectors. *Parasites and Vectors* 5: 21. [CrossRef]

Norkutė, Milda, Vasilis Sarafidis, Takashi Yamagata, and Guowei Cui. 2021. Instrumental Variable Estimation of Dynamic Linear Panel Data Models with Defactored Regressors and a Multifactor Error Structure. *Journal of Econometrics* 220: 416–46. [CrossRef]

Onatski, Alexei. 2010. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics* 92: 1004–16. [CrossRef]

Pesaran, M. Hashem. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74: 967–1012. [CrossRef]

Pesaran, M. Hashem. 2015. Testing Weak Cross-Sectional Dependence in Large Panels. *Econometric Reviews* 34: 1089–117. [CrossRef]

Pradhan, Amruta, and Veena Srinivasan. 2022. Do dams improve water security in India? A review of post facto assessments. *Water Security* 15: 100112. [CrossRef]

Sadoff, Claudia. 2018. *Dams Are a Breeding Ground for Mosquitoes—To Eradicate Malaria, We Must Rethink Their Design*. London: The Telegraph.

Sarsons, Heather. 2015. Rainfall and Conflict: A Cautionary Tale. *Journal of Development Economics* 115: 62–72. [CrossRef]

Sharma, V. P. 1998. Fighting malaria in India. *Current Science* 75: 1127–40.

Singh, Satyajit K. 1990. Evaluating Large Dams in India. *Economic and Political Weekly* 25: 561–74.

Skoufias, Emmanuel. 2003. Consumption smoothing in Russia. *Economics of Transition and Institutional Change* 11: 67–91. [CrossRef]

Strobl, Eric, and Robert O. Strobl. 2011. The distributional impact of large dams: Evidence from cropland productivity in Africa. *Journal of Development Economics* 96: 432–50. [CrossRef]

Wildi, Walter. 2010. Environmental hazards of dams and reservoirs. *Near Curriculum in Natural Environmental Science* 2: 187–97.

Yewhalaw, Delenasaw, Yehenew Getachew, Kora Tushune, Kifle W/Michael, Wondwossen Kassahun, Luc Duchateau, and Niko Speybroeck. 2013. The effect of dams and seasons on malaria incidence and anopheles abundance in Ethiopia. *BMC Infectious Diseases* 13: 161. [CrossRef] [PubMed]