POLLUTION EXTERNALITIES AND HEALTH: A STUDY OF INDIAN RIVERS

Samuel Stolper (Harvard)
joint with
Quy-Toan Do (The World Bank)
Shareen Joshi (Georgetown)

July 18th, 2014

ASIA PACIFIC | NYT NOW

Poor Sanitation in India May Afflict Well-Fed Children With Malnutrition

By GARDINER HARRIS JULY 13, 2014

"No Indian city has a comprehensive waste treatment system, and most Indian rivers are open sewers as a result"

"[Varanasi's] sewage plants can handle only about 20 percent of the sewage generated in the city"

"A stream of human waste – nearly 75 million liters per day – flows directly into the river just above the bathing ghats"

"Much of the city's drinking water comes from the river, and half of Indian households drink from contaminated supplies"

RESEARCH AGENDA

Two questions:

- What impact does water pollution in a district have on health outcomes in that district?
- Does the health impact persist downstream?

Context:

- Domestic water pollution in Indian rivers
- Infant mortality

Strategy:

- Use "upstream" water quality to construct an instrument
  - Compare IV (within-district externality) to reduced-form (downstream externality)

EPIDEMIOLOGY LITERATURE

Water use and mortality

- John Snow (1849) on the London cholera epidemic
  - The cause: Sewage in the form of fecal bacteria leaking into a public well
- Not just drinking water, but also irrigation (Cifuentes et al 2000), bathing, food, and person-to-person contact (Carr 2001)

Pathogens associated with domestic pollution

- Rotavirus; Campylobacter; Salmonella; Shigella; Adenovirus; E. Coli

Disease and illness associated with domestic pollution

- Cholera; Typhoid; Diarrhea; Hepatitis; Ascariasis; Schistosomiasis; Cryptosporidiosis

ECONOMICS LITERATURE

Water pollution and health: E.g., Kremer et al (2012), Spears (2012), Brainerd and Menon (2011)

Water pollution and India: Brandon and Homman (1995), Markyanda and Murty (2001), Dasgupta

Water pollution spillovers: Sigman (2002), Sigman (2005), Lipscomb and Mobarak (2007)

INDIAN CONTEXT

Central Pollution Control Board (2013) on sewage treatment:

- Total sewage burden, all cities >50K: 29,129 MLD
- Installed treatment capacity: 6,190 MLD
- Shortfall: 21,196 MLD (73% of total burden)

National River Conservation Plan (NRCP): 1986 - Present

- Water quality target: “Bathing class” standard
- “Interception, diversion, and treatment” of sewage
- US $80 billion spending in 3 decades
- No significant impact on river water quality (Greenstone and Hanna 2012)

METHODS

Data and Empirical Strategy

DATA OVERVIEW

Water pollution: Monitor-months from 1986-2004 (CPCB)

- 470 monitors along 162 rivers → 38,789 observations
- Latitude and longitude of monitors (authors' work)
- Use Fecal Coliforms (FCOLI) as our first-choice indicator of domestic pollution

Infant mortality: Two different samples of child-months

- DLHS-II, RCH: 264,375 births → 2.6 million observations
- DHS-II, NFHS: 39,125 births → 388,301 observations
- Children only identified by district, NOT town/village

Controls

- Climate (rainfall and air temperature)
- Birth, mother, and household characteristics

INFANT MORTALITY

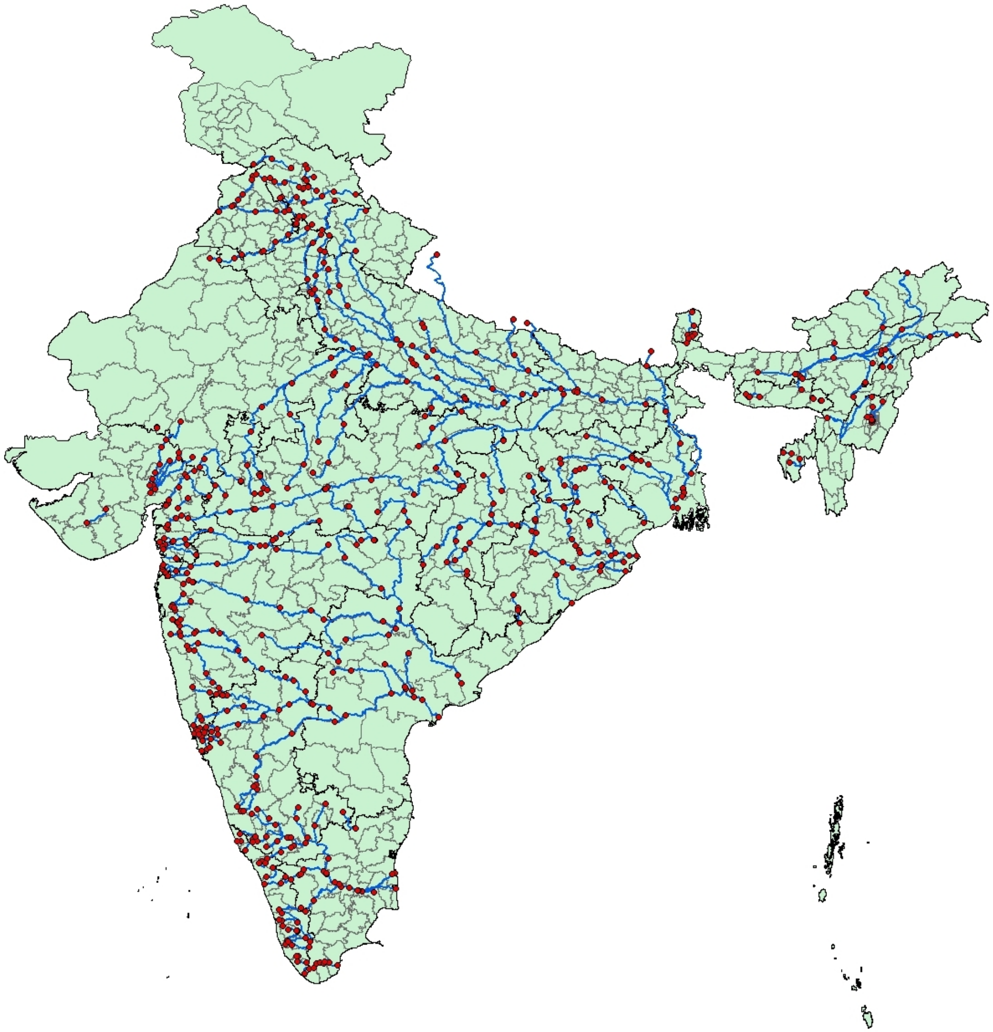
Table 3. Summary Statistics for Infants

		RCH-2 (1)	NFHS-2 (2)
1-Month Mortality	Mean	0.039	0.037
(0/1)	Std. Dev.	0.194	0.189
	N	264,375	39,125
1-Year Mortality	Mean	0.057	0.057
(0/1)	Std. Dev.	0.232	0.233
	N	264,375	39,125
1[Died this month]	Mean	0.006	0.006
(0/1)	Std. Dev.	0.077	0.08
	N	2,653,310	388,301

Notes

1. Column (1) provides statistics from the RCH-2 survey module, while Column (2) provides statistics from the NFHS-2 survey module.
2. All statistics are compiled from samples composed only of infants in districts for which there is matching pollution data.
3. Stats for the variable "1[Died this month]" are calculated across child-months; stats for the other two variables are calculated across children (not child-months).

MAPPING



INSTRUMENTAL VARIABLES (IV) STRATEGY

Challenge: Water pollution is not randomly assigned

- Locations with more water pollution may be more urban → Better health care or other infrastructure
- Locations with more water pollution may have more industrial or economic activity → Greater wealth

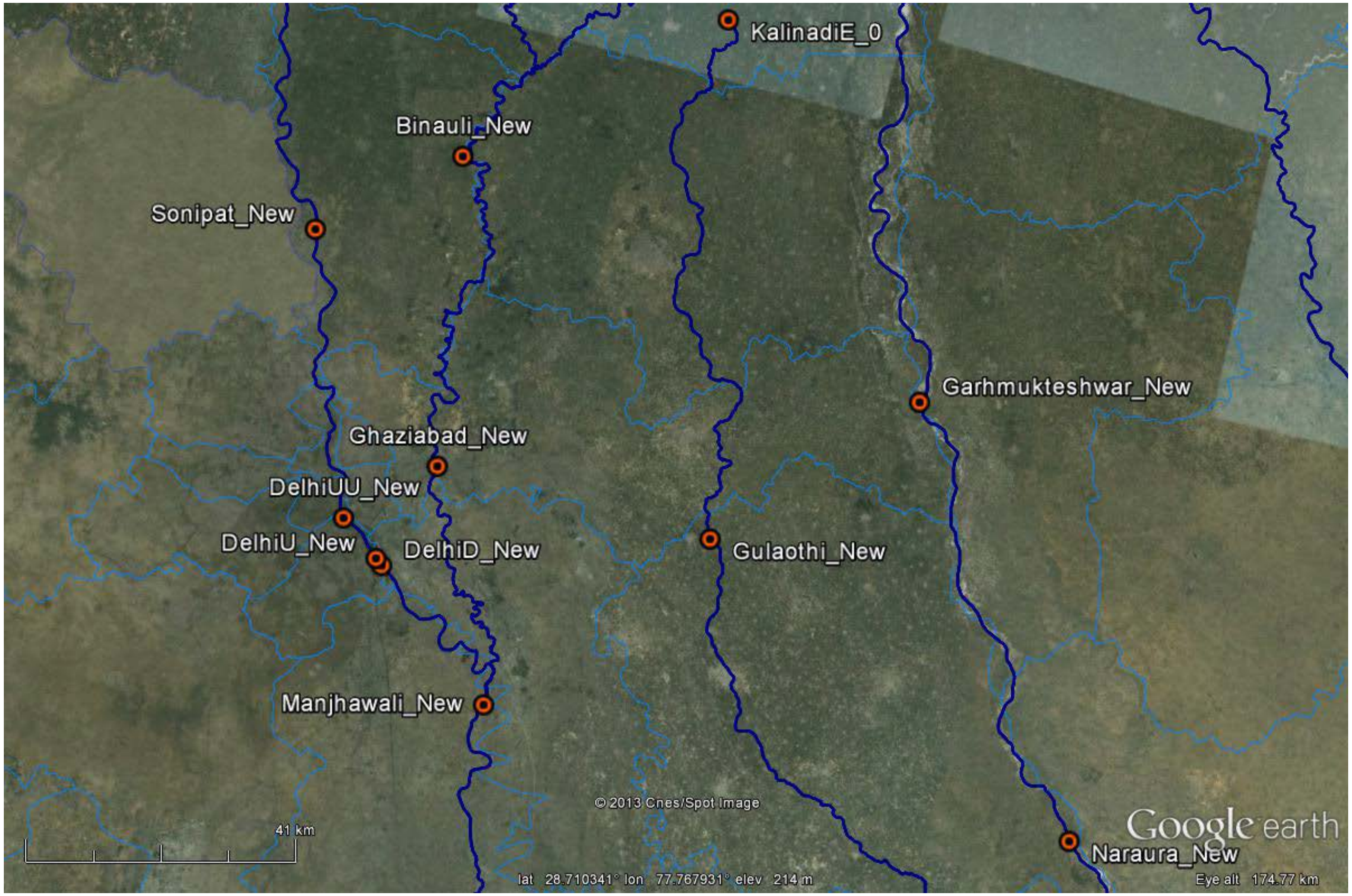
Risk of omitted variable bias: Spurious linkage between water pollution and health

- OLS: 'Naïve' estimate of pollution-health relationship

Solution: Isolate 'quasi-random' variation in pollution

- IV: First predict pollution in district X with pollution in the next district upstream of X; Then compare that prediction to infant mortality

UPSTREAM ASSIGNMENT



ECONOMETRICS

OLS (naïve, within-district)

$$IM_{dymi} = \beta_0 + \beta_1 P_{dym} + \beta_2 Rain_{dym} + \beta_3 Temp_{dym} + \beta_4 X + \lambda_d + \tau_{ym} + \phi_{sy} + \varepsilon_{dymi}$$

2SLS (within-district)

$$P_{dym} = \alpha_0 + \alpha_1 P_{(d+1)ym} + \alpha_2 Rain_{dym} + \alpha_3 Temp_{dym} + \alpha_4 X + \lambda_d + \tau_{ym} + \phi_{sy} + \varepsilon_{dymi}$$

$$IM_{dymi} = \beta_0 + \beta_1 \widehat{P_{dym}} + \beta_2 Rain_{dym} + \beta_3 Temp_{dym} + \beta_4 X + \lambda_d + \tau_{ym} + \phi_{sy} + \varepsilon_{dymi}$$

Reduced form (downstream)

$$IM_{dymi} = \beta_0 + \beta_1 P_{(d+1)ym} + \beta_2 Rain_{dym} + \beta_3 Temp_{dym} + \beta_4 X + \lambda_d + \tau_{ym} + \phi_{sy} + \varepsilon_{dymi}$$

RESULTS

First Stage and Full Two-Stage (Health) Regressions

FIRST-STAGE RESULTS

Table 5. First-Stage Pollution Regressions

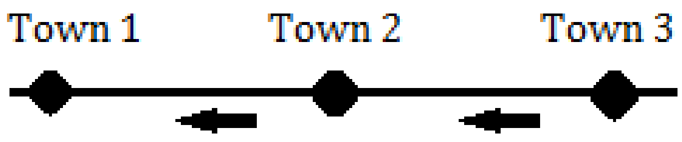
	MONITOR		DISTRICT
	(1)	(2)	(3)
Panel A. Bound of [50 km, 300 km]			
Upstream Log(FCOLI)	0.259***	0.358***	0.309***
	(0.051)	'(0.078)	(0.049)
Upstream Log(FCOLI)*Distance (km)		-0.0009*	
		(0.0005)	
R^2	0.81	0.81	0.83
N	55,700	55,700	19,900
Panel B. Bound of [75 km, 300 km]			
Log(FCOLI)	0.195***	0.387***	0.229***
	(0.053)	(0.097)	(0.047)
Upstream Log(FCOLI)*Distance (km)		-0.0014**	
		(0.0006)	
R^2	0.79	0.80	0.82
N	57,200	57,200	19,000

Notes

1. The dependent variable in all regressions is (downstream) Log(FCOLI).
2. All regressions include total rainfall and average air temperature as controls, as well as fixed effects by monitor-pair (or district) , year-month, and state-year.
3. All regressions cluster standard errors by monitor-pair.
4. Pollution values are computed as the moving average of the current month and the three previous ones.

TEST OF VALIDITY

Upstream pollution levels must predict downstream pollution levels only through the channel of pollution flow along the river



Once we predict pollution in Town 1 with pollution in Town 2, pollution in Town 3 should have no predictive power.

Mathematically:

- $Cov(P_{downstream}, P_{2nd\text{-}upstream}) > 0$
- $Cov(P_{downstream}, P_{2nd\text{-}upstream} \mid P_{1st\text{-}upstream}) = 0$

TEST OF VALIDITY - RESULTS

Table 6. First-Stage Validity Test Results

	(1)	(2)	(3)
Panel A. Bound of [50 km, 300 km]			
1st-Upstream Log(FCOLI)	0.200***		0.197***
	(0.048)		(0.048)
2nd-Upstream Log(FCOLI)		0.045**	0.019
		(0.020)	(0.015)
R^2	0.84	0.84	0.84
N	30,400	30,400	30,400
Panel B. Bound of [75 km, 300 km]			
1st-Upstream Log(FCOLI)	0.150***		0.149***
	(0.051)		(0.051)
2nd-Upstream Log(FCOLI)		0.020*	0.010
		(0.012)	(0.012)
R^2	0.83	0.83	0.83
N	28,900	28,900	28,900

Notes

1. The sample is restricted to all monitor-months for which there is pollution data at downstream, 1st-upstream, and 2nd-upstream locations.
2. The dependent variable in all regressions is (downstream) Log(FCOLI).
3. All regressions include total rainfall and average air temperature as controls, as well as fixed effects by monitor-pair (or district) , year-month, and state-year.
4. All regressions cluster standard errors at the monitor-pair level.
5. Pollution values are computed as the moving average of the current month and the three previous ones.

HEALTH REGRESSIONS – RCH-2 RESULTS

Table 7. Health Regression Results, RCH-2

	OLS		IV		RF	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Bound of [50 km, 300 km]						
Log(FCOLI)	-0.065	-0.07	0.041	-0.291	0.052	-0.057
	(.)	(.)	(0.194)	(0.248)	(0.051)	(0.055)
Log(FCOLI)*1[Age=1 month]		0.051		3.676***		1.272***
		(.)		(1.158)		(0.376)
Implied elasticity for one-month-olds				*0.09*		*0.03*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	1,870,000	1,870,000	1,790,000	1,790,000	2,190,000	2,190,000
Panel B. Bound of [75 km, 300 km]						
Log(FCOLI)	-0.08	-0.102	0.17	-0.088	0.062	-0.045
	(.)	(.)	(0.280)	(0.281)	(0.052)	(0.056)
Log(FCOLI)*1[Age=1 month]		0.254		3.557***		1.256***
		(.)		(1.143)		(0.368)
Implied elasticity for one-month-olds				*0.09*		*0.03*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	1,780,000	1,780,000	1,700,000	1,700,000	2,100,000	2,100,000

Notes
1. An observation is a child-month.
2. Pollution and weather values are aggregated to the district-level for matching with infants.
3. The dependent variable in all regressions is 1[Died this month].
4. All regressions include fixed effects at the year-month, state-district, and state-year levels and the full set of controls.
5. All regressions cluster standard errors at the district level and use sampling weights provided by DLHS-2.
6. Pollution values are computed as the moving average of the current month and the three previous ones.

HEALTH REGRESSIONS – NFHS-2 RESULTS

Table 8. Health Regression Results, NFHS-2

	OLS		IV		RF	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Bound of [50 km, 300 km]						
Log(FCOLI)	-0.066	-0.174	-1.802	-0.345	-0.042	-0.193
	(.)	(.)	(158.219)	(0.464)	(0.126)	(0.133)
Log(FCOLI)*1[Age=1 month]		1.362		4.755***		1.755***
		(.)		(1.616)		(0.586)
Implied elasticity for one-month-olds				*0.13*		*0.05*
R^2	0.03	0.03	0.01	0.03	0.03	0.03
N	280,000	280,000	271,000	271,000	325,000	325,000
Panel B. Bound of [75 km, 300 km]						
Log(FCOLI)	-0.028	-0.152	0.113	-0.315	-0.006	-0.163
	(0.149)	(0.160)	(0.650)	(0.646)	(0.126)	(0.125)
Log(FCOLI)*1[Age=1 month]		1.485**		5.057***		1.822***
		(0.659)		(1.752)		(0.589)
Implied elasticity for one-month-olds				*0.14*		*0.05*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	271,000	271,000	259,000	259,000	313,000	313,000

Notes
1. An observation is a child-month.
2. Pollution and weather values are aggregated to the district-level for matching with infants.
3. The dependent variable in all regressions is 1[Died this month].
4. All regressions include fixed effects at the year-month, state-district, and state-year levels and the full set of controls.
5. All regressions cluster standard errors at the district level.
6. Pollution values are computed as the moving average of the current month and the three previous ones.

DISCUSSION POINTS

Details of the Water Pollution–Health Relationship

REVIEW

What we've found:

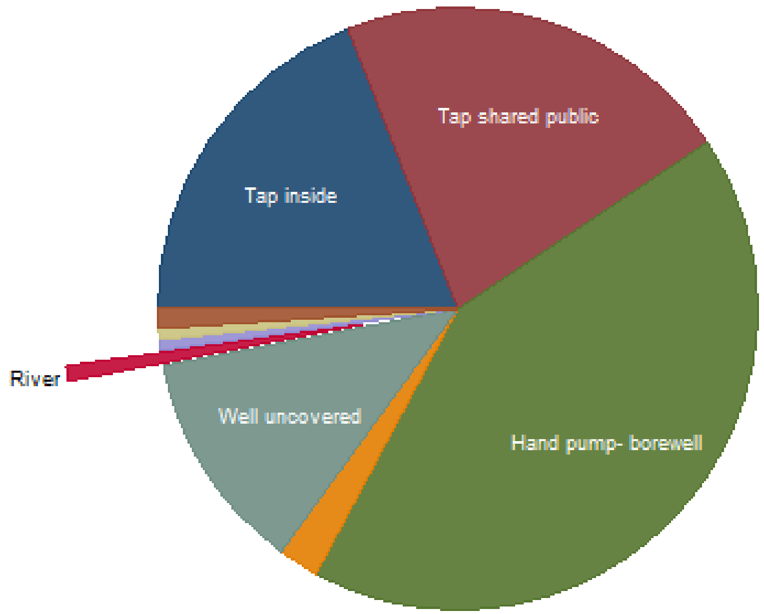
- Domestic river pollution appears to have a strong negative impact on survival in the first month, but not thereafter
  - Provides a causal estimate of the direct water pollution – health relationship
- The mortality burden of river pollution persists at districts downstream of pollution measurement
  - Links regional pollution spillovers to actual external (health) costs

Caveats

- Measurement error prevents us from having precise estimates of the water pollution – mortality relationship
  - We do not know where infants live
  - We do not know whether or how infants interact with river water
- Our research cannot, as yet, suggest the precise mechanism by which water pollution kills infants
  - Drinking, bathing, eating, irrigation, and person-to-person contact are all possible channels

DRINKING WATER

Figure 2. Primary Drinking Water Source, RCH-2



Notes

1. The counts above are taken from all births, rather than all households.

LINKING TO POLICY

- Despite economic growth, improvements in sanitation, expanded access to piped water, and long-term policies to clean up rivers, there remains high river-pollution related infant mortality
- Even without making causal claims on what behavior drives this mortality burden, our findings suggest that we need to do more to clean up rivers
- The evolution of our estimated relationship *over time*, in comparison with trends in piped water access, sanitation, and baseline water quality, may aid a discussion of what policies are best

NEXT STEPS

- Model the evolution of mortality impacts over time
- Compile summary data on time trends in piped water access, sanitation coverage and capacity, baseline water quality, baseline perinatal care
- Investigate the impacts of prenatal (fetal) exposure to water pollution