Identifying the Hidden Costs of a Public Health Success:

Arsenic Well Water Contamination and Productivity in Bangladesh

Mark Pitt, Mark Rosenzweig and Nazmul Hassan

.

This paper provides estimates of the impact of arsenic retention in the body on

> cognitive performance, schooling, strength, body size, conventional measures of morbidity, earnings, household consumption and household productivity by gender and age

Important issue in Bangladesh where 57 million people are exposed to arsenic levels above WHO standards by drinking, irrigating and cooking with contaminated water

Result of a major public health intervention:

> To reduce diarrheal disease, the government of Bangladesh, with international support, promoted the digging of wells to provide "clean" water
>
> A success: Shift from surface water sources to groundwater (95% of the population relies on well water) reduced morbidity from diarrheal disease substantially - measured increases in body mass and stature for all
>
> But, it turned out in 59 of 64 districts the well water was contaminated by naturally-occurring arsenic
>
> *This is the largest poisoning of a population in the history of the world*

The link between arsenic exposure and retention has two components:

A. Ingestion: drinking contaminated water

consuming foods cooked with contaminated water, irrigated with contaminated water

B. Retention: the body metabolizes the arsenic and secretes ingested arsenic (process is methylation)

Diet:

Evidence suggests that dietary sources of arsenic are the most important source of ingestion in Bangladesh (variation in diet for given exposure)

Foods differ in the degree to which they absorb arsenic from irrigation and cooking water (ingestion) and aid, from specific nutrients like folate, in methylation

Leafy vegetables have the highest retention of arsenic and also high in folates - net effect unknown

Genetics:

People differ in the degree to which they metabolize arsenic (specific genes)

Ability to methylate is correlated across kin

Genes linked to methylation not correlated with genes linked to cognition ability, strength (we show)

Existing evidence:

Does not take into account endogeneity of diet choices, well proximity

Arsenic retention is importantly a matter of choice

Small samples, so no ability to assess whether the effects differ by age or gender

Little evidence on capability, productivity, economic effects or outcomes

Mostly skin lesions, rare cancers, mortality

The consequences of water without organic contaminants very visible:

A. Rise in body mass at all ages, for men and women, with no increase in calorie consumption

B. Rise in adult height

C. Rise in schooling attainment, attendance, especially for women in part due to better health

The consequences of arsenic poisoning less clear

Most of the population is unaware of their retained arsenic, none know the costs

Cannot perform a randomized trial measuring the direct effects of arsenic poisoning

Productivity consequences are also so far invisible to the research community

Figure A. Height (cm) by Year Attained Age 22 and Gender, 1970 – 2004 for Respondents Aged 22-59 in 2007-8 (Lowess-smoothed)

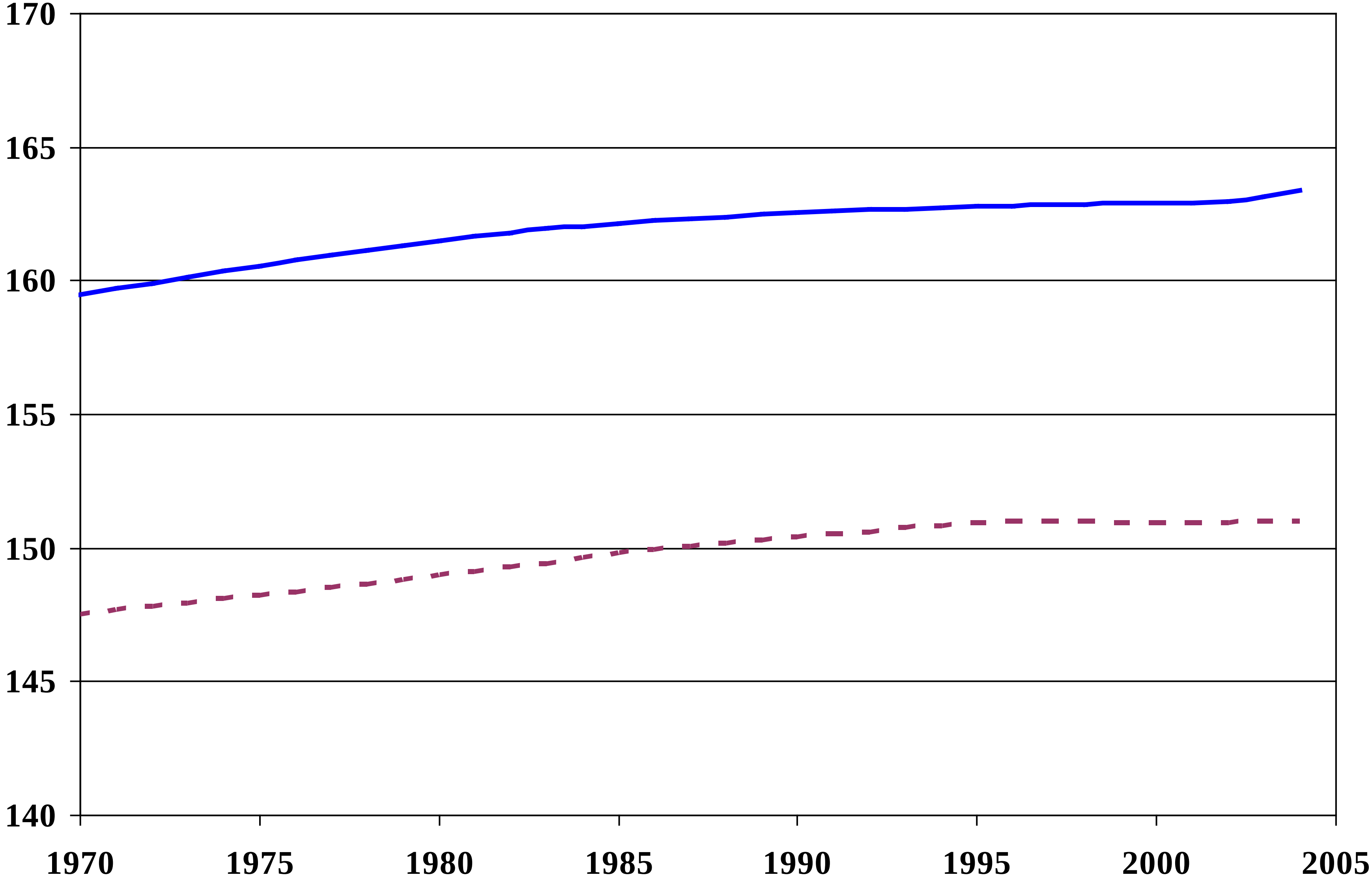
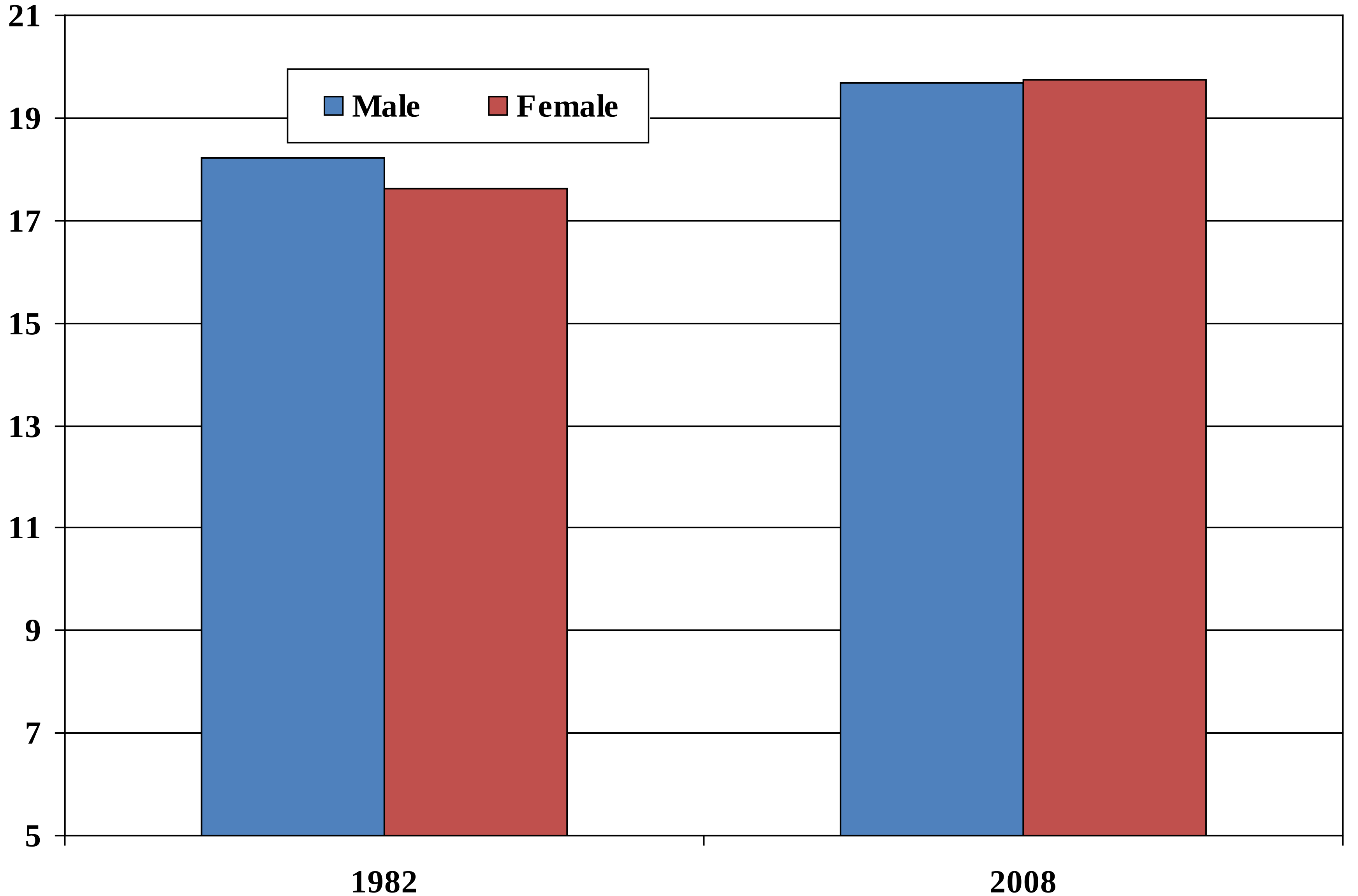


Figure B. BMI in 1981-82 and 2007-8, by Gender



We use new panel data and a well-established biomarker for short-term arsenic retention in the body based on toenail clippings to obtain causal estimates of

A. The effects of different food groups on arsenic retention

B. The effects of arsenic retention, by gender and age, on

1. Direct measures of capabilities: cognition, strength, by gender

2. Human capital outcomes: schooling, anthropometrics

3. Choice of occupation, entrepreneurship, labor supply

4. Earnings (males), household productivity (females) and household consumption

Estimation procedure exploits

A. Genetic variation and within-family correlation in abilities to methylate arsenic

B. Spatial separation of family members due to migration (e.g., marriage)

Important challenges for measuring economic consequences:

1. Most men do not work for wages, or work for wages only part of the year

Looking at only wage effects is selective, can be misleading (arsenic poisoning can affect who works in what occupations)

2. Most male earnings are from self-employment, where other factors of production matter

3. Few women are in the labor market

4. But, women contribute importantly to home production

No assessment of economic consequences of any health problem or treatment should ignore this important aspect of production in Bangladesh: household productivity

Modeling the Effects of Arsenic Contamination

Begin with one-person model to illustrate how

A. water quality, preferences, health information, and abilities to methylate ingested arsenic

B. optimizing behavior

affect inferences about the relationship between an individual's measured amount of retained arsenic and measures of capabilities and economic outcomes.

Basic technological relationship

(1) $$A_{ij} = \mathrm{A}(\boldsymbol{C_{ijl}}, \omega_{ij}, \mu_{ij})$$

where

A_{ij} = retained arsenic for individual *I* residing in environment *j*

$\boldsymbol{C_{ijl}}$ = vector of consumed foods indexed by *l*;

ω_{ij} = the quality (lower arsenic content) of the water used for cooking and drinking;

μ_{ij} = the individual's endowed ability to methylate arsenic.

Consumed water quality depends on the quality of the water source e_j and the individual's purification effort t_{ij}:

(2) $$\omega_{ij} = \omega(t_{ij}) + e_j ,$$

where $\omega' > 0$, $\omega'' < 0$.

What is t_{ij}?

may include time spent fetching water from alternative sources, time boiling or otherwise treating water from water sources that have biological contaminants (but lower arsenic and/or time spent collecting additional fuel needed for water boiling.

The budget constraint is given by

(3) $$F_{ij} + (\Omega - t_{ij})w_{ij} = \Sigma p_{jl}C_{ijl},$$

where F_{ij}=non-earnings sources of income, the p_{jl} = local food prices.

$w_{ij} = w_j h_{ij}$, where

h_{ij} = the individual's capability or skill h_{ij},

w_j = local per-unit rental price of skill

Skill is affected negatively by retained arsenic and positively by a skill endowment r_{ij}:

(4) $$h_{ij} = (A_{ij}, r_{ij}),$$

where $h_1<0$.

The association between measures of individual-specific retained arsenic A_{ij} and, say, productivity h_{ij} in a given environment (given water quality e_j and local prices p_{jl}) in j is:

(5) $$dh_{ij}/dA_{ij} = h_A(1 + A_{2j}\omega'(dt_{ij}/d\mu_{ij})(d\mu_{ij}/dA_{ij}) + A_1(dC_{ij}/dF_{ij})(dF_{ij}/dh_{ij})),$$

Which is NOT the same as the causal effect of retained arsenic of productivity (h_A)

Similarly, the association between measures of the quality of water consumed and A_{ij} also reflects behavior, and not just the technological relationship (A_2):

(6) $$dA_{ij}/d\omega_{ij} = A_2 + h_A w_j \omega' dt_{ij}/dw_{ij}$$

Is it plausible that arsenic-contaminated water affects behavior if the costs, the amounts of arsenic ingested or retained, and productivity effects are unknown?

Assume that agents only have *public health* information:

(I) they know that arsenic is deleterious to health,

(ii) they know which non-local water sources reduce arsenic ingestion

(iii) they know how water purification effort affects arsenic retention.

However, they are *un*informed about

(I) the effects of *A*s on productivity *h*,

(ii) their own individual arsenic retention A_{ij}

(iii) their own methylation efficiency endowment μ_{ij}.

That is, the *individual* effects of arsenic retention are “hidden”.

Assume people dislike arsenic and like to consume; the utility function is

(7) $$U = U(A_{ij}^*, C_{ij}; u_{ij}) \quad U_1<0,\ U_2>0$$

where A_{ij}^* = the agent’s beliefs about his retained arsenic (u_{ij} = preferences)

What is the optimal time spent fetching water?

The FOC is:

(8) $$U_A A_2{}^* \omega' = U_c w_{ij},$$

Agents face a trade-off between consumption (good) and arsenic (bad) because expending effort to increase water quality reduces the time available for income earning.

Note that if agents were informed about the economic consequences of arsenic retention the FOC is:

(9) $$U_A A_2{}^* \omega' = U_C(w_{ij} - A_2{}^* \omega' h_1 w_j(\Omega - t_{ij}));$$

Comparing (8) to (9) indicates that

> lack of knowledge about the relationship between arsenic retention and productivity in the population leads to higher levels of contaminated water consumption (less water purification fort) than is optimal.

There is a productivity and health payoff to the dissemination of information about the economic consequences of arsenic retention, if there are such consequences.

Can also be proved that, even in this simple model, given realistic information constraints, the bias in the association between individual arsenic retention and productivity cannot be signed.

One cannot therefore use observations on arsenic retention and productivity to infer causation from arsenic to productivity.

What about using the relationship between water quality and, income ($dF_j/de_j = dC_j/de_j$) to infer the effects of arsenic retention on productivity (h_1)? *No*

Improvements of the local water source (change in e_j) also affect behavior:

(11) $$dt_{ij}/de_j = \omega'[(A_2)^2 U_{AA} + U_A A_{22}]/\Phi + A_2 h_1[U_C w_j/\Phi + (\Omega - t_{ij})w_j\, dt_{ij}/dF_j]$$

The income effect inclusive of the behavioral response is

(12) $$dC_j/de_j = (\Omega - t_{ij})w_j h_1 A_2(1 + \omega'(dt_{ij}/de_j)) - (dt_{ij}/de_j)w_{ij}\ ,$$

which differs from the productivity effect h_1 by $-(dt_{ij}/de_j)[(\Omega - t_{ij})w_j h_1 A_2 \omega' + w_{ij}]$.

As retention in a multiple-member household

The above simple model is not well-suited for inferring the economic consequences of arsenic ingestion and retention in Bangladesh:

A. In Bangladesh time devoted to activities associated with consumed water quality are typically done by non-earning women.

In our 2008 data, 98% of household members who spend any time fetching water, gathering fuel (wood or dung), and/or cooking are women (80% are wives of the head). Less than 3% also participated in market or household earning activities.

B. Earners are almost exclusively men, and if there are multiple men they are usually kin (brothers and/or sons of the household head).

C. A large fraction of earning men, as is typical in many low-income countries, are also self-employed, making it difficult to directly measure earnings strictly associated with individual capabilities or skills.

59.7% of men in the labor force aged 24-59 are primarily self-employed in our 2008 data.

We show in a model that incorporates these features of rural Bangladesh, and endogenous water purification effort, that it is possible to identify:

A. The existence of any home productivity effects of arsenic retention

B. The magnitude of market productivity effects

with information on exogenous variation in individual-specific arsenic retention.

We assume that a household consists of N identical earners plus one non-earner (wife) who produces the home good *M* and also spends time in water purification activities.

We add a production function for the home good (13) and a home productivity *H* or skill function for the home producer (14), where the *f* subscript refers to the woman engaged in home production:

(13) $$M = \mathrm{M}(\Omega - t_{fj}, H_{fj}) \quad \mathrm{M}_1 > 0, \mathrm{M}_2 > 0, \mathrm{M}_{12} > 0$$

(14) $$H_{fj} = \mathrm{H}(A_{fj}, r_{ij}) \qquad \mathrm{H}_1 < 0. \quad \text{(Home productivity)}$$

We assume, in accord with the genetics literature, that the methylation efficiency endowment has a common family component, so that $\mu_{ij} = \mu_j + \xi_{ij}$ for the earners.

The utility function for the multi-member household is

(15) $$U = (A_j, C_j, M_j) \qquad U_1<0;\ U_2,U_3>0;\ U_{11}>0;\ U_{22},U_{33}<0,$$

where A_j=average arsenic retention of family members.

The budget constraint is given by

(16) $$F + N\Omega h_{ij} w_j = C_j.$$

In this model, the spouse's retention of arsenic has no direct effect on earned income or the amount of the consumption good purchased.

Whether or not there are home productivity effects of arsenic, however, there is still a trade-off between the home-produced good and arsenic retention in the household.

Given the same informational assumptions as in the first model the FOC is

(17) $$U_A A_2 \omega' = U_M M_1$$

With information on the exogenous variation in arsenic retention among earners and non-earners, the model indicates that it is possible to

1. Quantify the effects of retained arsenic on market productivity (h_1) using the relationship between retained arsenic among earners and total household expenditures. This is because

$$dC_j / dA_{Ej} = N w_j \Omega h_1$$

2. Identify the existence of home productivity effects of arsenic retention among nonearners.

This is because the model delivers the following proposition for identifying the existence of effects of (the nonearning wife's) arsenic retention on home productivity:

> *If H* (*skill*) *does not matter in producing the home good or arsenic does not reduce home productivity, the wife's individual arsenic retention will have no effect on water purification effort* (*e.g., changing water sources*).

3. Infer the relationship between home-produced and purchased goods.

> This is because the model indicates that: *Lower arsenic retention among household earners will affect water purification effort by a non-earning household member only if retained arsenic affects market productivity and if the home good and purchased goods are not separable in* (15).

Proofs:

1. If, as assumed before, productivity effects of arsenic are not known, an increase in the wife's methylation ability on her effort to improve water is given by:

(18) $$dt_{fj}/d\mu_{fj} = \{A_3H_1[U_{MM}M_1M_2 + U_MM_{12}]/\Phi^*,$$

where $\Phi^*<0$. If M_2 and $M_{12} = 0$, expression (18) vanishes. If household productivity rises as arsenic retention falls, lowering arsenic retention will reduce effort to improve water quality.

2. The common effect of a reduction in arsenic among the earners on the wife's effort is:

(19) $$\mathrm{d}t_{fj}/\mathrm{d}\mu_j = \mathrm{N}w_j\Omega \mathrm{h}_1\mathrm{A}_3\mathrm{U}_{MC}\,\mathrm{M}_1/\Phi^* = \mathrm{N}w_j\Omega \mathrm{h}_1\mathrm{A}_3\mathrm{d}t_{ij}/\mathrm{d}F_j,$$

Expression (19) vanishes if *M* and *C* are separable; reducing arsenic retention among the earners increases (decreases) t_{fj} if *C* and *M* are complements, $\mathrm{U}_{MC}>0$ (substitutes, $\mathrm{U}_{MC}<0$).

Identifying arsenic retention effects using panel data

Our first objective is to estimate the effect of *As* on measures of individual productivity and health.

Linearizing the productivity function (4), we seek to identify the parameter δ in:

(21) $$h_{ijl} = \delta A_{\boldsymbol{ijl}} + \boldsymbol{Z}_{ijl}\boldsymbol{\beta}_z + u_{ijl} + u_l + \varepsilon_{ijl},$$

where the l index identifies the individual's relationship with a family member,

$\boldsymbol{Z}_{ijl}$ = a vector of observed exogenous attributes of the individual and household,

u_{ijl} = an individual-specific error, u_j = a household fixed effect (reflecting, for example, the local health environment), and ε_{ijl} an iid error.

A_{ijl} is correlated with the error terms containing, for example, preferences for foods, area-level prices and local water quality.

Thus least squares estimation of (21) would not provide a consistent estimate of δ.

Randomized interventions are not helpful as instruments:

A. At the village level, such as the construction of a deep well to improve e_j, cannot be used because the new well will directly alter the allocation of time of all family members and thus affect labor market and other outcomes directly for any individual.

B. At the individual level, such as interventions that reduce arsenic ingestion or decrease arsenic retention through nutrition (folates), have direct effects on productivity.

What is required is exogenous variation in a variable that directly affects arsenic retained in the body, given a person's exposure to arsenic in the environment, and has no other direct effects on the outcomes of interest.

Our strategy for the identification of δ exploits exogenous individual variation in genes that influence an individual's ability to metabolize arsenic, as embodied in μ_{ijl} in the model, and the consequent genetic linkages among family members in that ability.

Recent evidence indicates that genetic variations (polymorphisms) are a major source of exogenous variation in the arsenic stock of the body within the same exposure area (Vahter, 2000).

Methylation of arsenic facilitates its excretion from the body. Arsenic is transformed in the body, and the end-products of the methylation process are metabolites - MMA (monomethylarsonic acid) and DMA (dimethylarsinic acid) - that are readily excreted. The enzymes that are required for this chemical process have also been identified and linked to four specific genes:

1. Arsenite methyltransferase catalyzes the oxidative methylation of arsenic to forms of MMA. This enzyme is encoded on a gene called AS3MT on human chromosome 10.

2. The glutathione S-transferase omega-1 variant of the glutathione-S-transferase (GST) enzyme is encoded by the GSTO1 gene, located on human chromosome 10.

3. The glutathione S-transferase theta-1 variant is encoded by the GSTT1 gene, located on human chromosome 10.

4. Methylenetetrahydrofolate reductase (MTHFR gene).

Example of evidence of the effects of gene variation on arsenic symptoms:

Ahsan *et al.* (2007) estimate that the proportion of skin lesions in their study population in Bangladesh that is attributable to polymorphisms in the MTHFR gene is 7.5 percent, and the proportion due to polymorphisms in the GSTO1 gene is 8.9 percent. Steinmaus et al. (2007) find that polymorphisms in the MTHFR gene are associated with arsenic methylation efficiency in an Argentinian sample.

The genetic origins of arsenic metabolism suggest that the ability to methylate is correlated among family members.

Chung *et al.* (2002) based on families from Chile with long-term exposure to very high levels of arsenic in drinking water (735–762 µg/L) found that 13–52% of the variations in the methylation patterns were from being a member of a specific family.

In our data, the correlation between the *As* of family members measured from the collected toenail clippings also appear to reflect genetic origins - the correlation in the measured *As* concentrations of heads and wives (who are only in few cases even distant relatives), net of a village fixed effect, is a statistically significant 17.3% lower than the correlation between those of heads and their co-resident mothers in the same set of households.

We use as an instrument for an individual's retained arsenic an estimate of the individual's genetic ability to metabolize (methylate) arsenic based on a non co-resident family member's ability to do so net of common factors in the environment.

We do not have DNA information, so we must employ indirect methods.

Using the notation of the model, the covariance between measured arsenic retention of two family members residing in areas *m* and *n* is

(22) $$\mathrm{cov}(A_{ijml}, A_{kjnl}) = \mathrm{var}(u_l)\mathrm{A}_3 + \mathrm{cov}(e_m, e_n)\mathrm{A}_2 + \mathrm{cov}(w_m, w_n)\mathrm{A}_2\omega'\mathrm{d}t/\mathrm{d}w + \mathrm{cov}(u_{ijl}, u_{kjn})\mathrm{A}_1\mathrm{d}C/\mathrm{d}u + \mathrm{cov}(P_m, P_n)\mathrm{d}C/\mathrm{d}P,$$

where μ_l is the common family genetic component of μ_{ijl}.

See that as long as local prices, local wages and the quality of water sources are spatially correlated, and if preferences among family members for foods are also correlated even if living apart, using the actual arsenic levels among spatially separated kin as an instrument for an individual's retained arsenic would NOT be appropriate for identifying δ.

The *As* covariation reflects both commonality of environments and common behavior.

We need to isolate that component of A_{ijml} that is unexplained by exposure to either environmental arsenic or by endogenously-determined individual-specific nutritional intakes but that contains the genetic component of methylation ability u_l.

To remove those components that are endogenously-determined through food and water choice and that reflect common environmental sources of arsenic from the measured arsenic, we first estimate the individual-specific nutrition production function for arsenic concentrations (1) using the measures of A_{ijml} taken from the toenail clippings.

Assuming a Cobb-Douglas form for that function, and taking logs, the equation we estimate is

(23) $$\text{Log}A_{ijk} = \boldsymbol{N}_{ijk}\boldsymbol{\alpha} + \mu_l + \xi_{ijk} + e_k + u_j + v_{ijk},$$

where $\boldsymbol{N}_{ijk}$ is a vector of person-specific and endogenous family inputs.

The inputs include the log of individual foods consumed, the log of cigarettes smoked, and the household's choice of a water source for cooking;

Equation (23) also contains five sources of unobserved heterogeneity highlighted in the model:

1. The genetic component of arsenic methylation ability that is shared among a lineage or kin group μ_l,
2. The individual-specific component of methylation ability.
3. Unmeasured exposure to environmental arsenic from local drinking and cooking water, e_k.
4. A household error component u_j
5. An iid error term v_{ijk}.

To obtain consistent estimates of the input-coefficient vector $\boldsymbol{\alpha}$, the effects of variation in diet and other behaviors on retained arsenic, we need to take into account that food intakes are affected by variation in these unobservables even if respondents are unaware of either e_k or μ_{ijk} (because, as we have shown, both may affect incomes if arsenic retention affects productivity).

To control for environmental water quality, we include in (23) a complete set of village dummy variables.

To deal with the correlation between the $\boldsymbol{N}_{ijk}$ and u_j and μ_{ijk} we estimate (23) by instrumental variables.

The instruments are: village-level prices of foods P_{kl} interacted with exogenous person- and household-specific exogenous attributes (own age and gender, the household head's age and the household value of landholdings (F_j).

Estimation of the nutrient effects on individual-specific arsenic concentrations thus exploits the real variation in relative food prices across the large number of villages in our sample, the existence of extended families in Bangladesh, and the information in our data containing individual food intakes.

For the estimates of the effects of food and water sources on arsenic concentrations to be credible based on the short-period information on food intakes, it is important that the measure of individual arsenic concentrations, based on the toenail assays, reflects relatively recent, and not lifetime, accumulation.

The toenail-based measure mainly reflects arsenic ingestion and excretion in recent months, not years (Kile *et al.*, 2005).

The residuals from (23) for genetically-linked but non co-resident family members are used as instruments for $A_{ij\mathrm{k}}$ in (21) to identify δ.

These residuals contain the genetic component of arsenic μ_{ijkl}, plus any measurement error, and the household fixed effect u_j.

By using residuals from non coresident family members who have resided in a different village at for least one year we minimize the influence of the household component.

We exploit the panel design of the survey, which followed all household members who left the households in a prior round, and make use of the fact that for almost every household in our original 1982 sample a relative had left the household between 1982 and the second round of the survey in 2002.

> *Household division is mostly due to marriage - between 1982 and 2002, for example, 85% of girls age 2-14 in 1982 had left their original household and village and 10% of the boys left the village.*

After estimation of (23), we compute from the estimated person-specific residuals the expectation $E(\mu_{ln}|TN_{ijk}$ - $\boldsymbol{N}_{ijk}\boldsymbol{\alpha}$ for all members of lineage *l* except person *ijl*).

The covariation between a non-coresident family member's μ_{ln} and respondent arsenic retention A_{ijkl} is thus $A_3 \text{var}(u_{ln})$.

The reduced-form covariation between a non-coresident family member's μ_{ln} and respondent productivity h_{ijkl} is:

$$\mathrm{Cov}(\mu_{ln}, h_{ijkl}) = (\mathrm{d}h_{ij}/\mathrm{d}\mu_{ij})\mathrm{var}(u_{ln}) + (\mathrm{d}h_{ij}/\mathrm{d}r_{ij})\mathrm{cov}(r_{ij}, \mu_{ln}).$$

Thus, there are two requirements for μ_{kn} to be a valid instrument for A_{ijk}:

1. The variance of the common genetic component of arsenic methylation ($\mathrm{var}(u_{ln})$) must be nontrivial - strong family links.

2. The $\mathrm{cov}(r_{ij}, \mu_{kn})$ must be negligible, where r_{ij}= genetic component of i's productivity - the genetic polymorphisms that regulate the efficiency of arsenic metabolism are unrelated to genes that affect the outcomes of interest h_{ijk}.

Testing for orthogonality among sources of genetic variation is critical.

We will use new data from a sample of genetically-linked individuals whose genome has been well characterized as part of the International HapMap Project.

> The method requires the determination of a set of loci on the human genome associated with (i) arsenic methylation and with (ii) human capital and productivity outcomes (general intelligence, body mass and height, and muscle development and strength).
>
> The location of these sets of genes suggests that they are unlikely to be correlated. However, we will obtain these correlations and test for their statistical significance.

The Survey Data

1. Survey design

   A. Panel survey tracking all individuals originally residing in 14 villages in rural Bangladesh in 1982

      3% attrition in 2002 round

   B. Added random sample of households in 2002 round, unrelated to original households

   C. 2007/8 round: followed all persons in 2002 round

2. Results: information on many separated kin

A. By 2002, 85% of women under age 15 had left the origin village; 10% of men in the same age group

B. In the 2007/8 round, households resided in 612 villages (from the original 14!)

3. Information on anthropometrics, individual-specific food intakes, tests

4. Measurement of individual-specific arsenic concentrations A_s

Based on clippings from all ten toenails for respondents aged 8 and over (7,356 individuals)

Preferred biomarkers for arsenic retention when exposure lasts for more than a few months (three) and significant share of exposure comes from food consumption

Requires trace metal analysis using inductively-coupled plasma mass spectrometry

Lab analysis expensive, so confined tests to:

Pairs of households containing related kin residing in different villages (N = 4,260)

Sample over-represents married women, aged (schooling, landholdings distributions not different)

Sample: **17X** the average concentration of A_s compared to control (US graduate students (N= 25)); also more dispersed - double the CV

Figure 1. Distribution of *As* Concentrations (ppb), by Sample

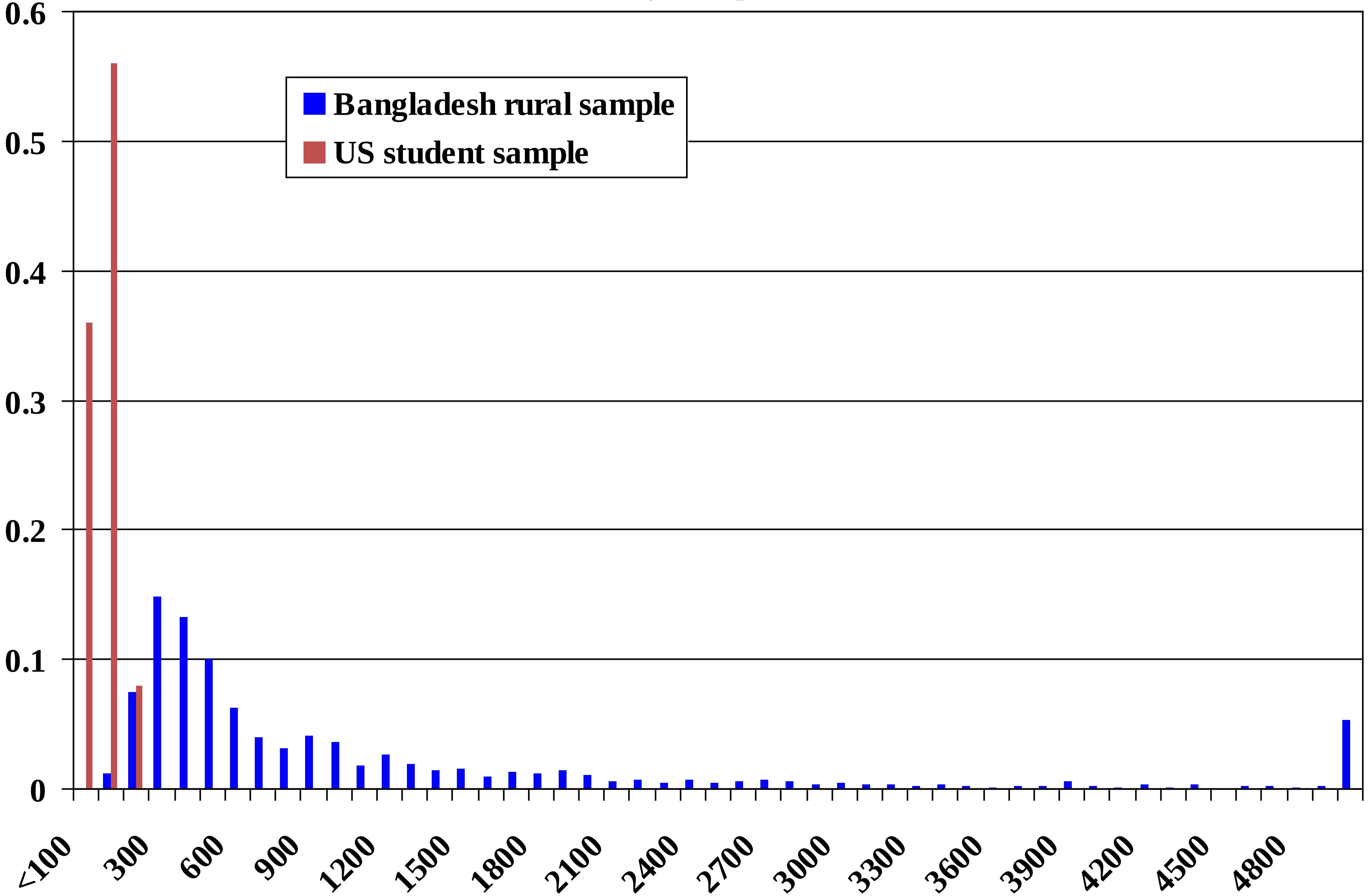
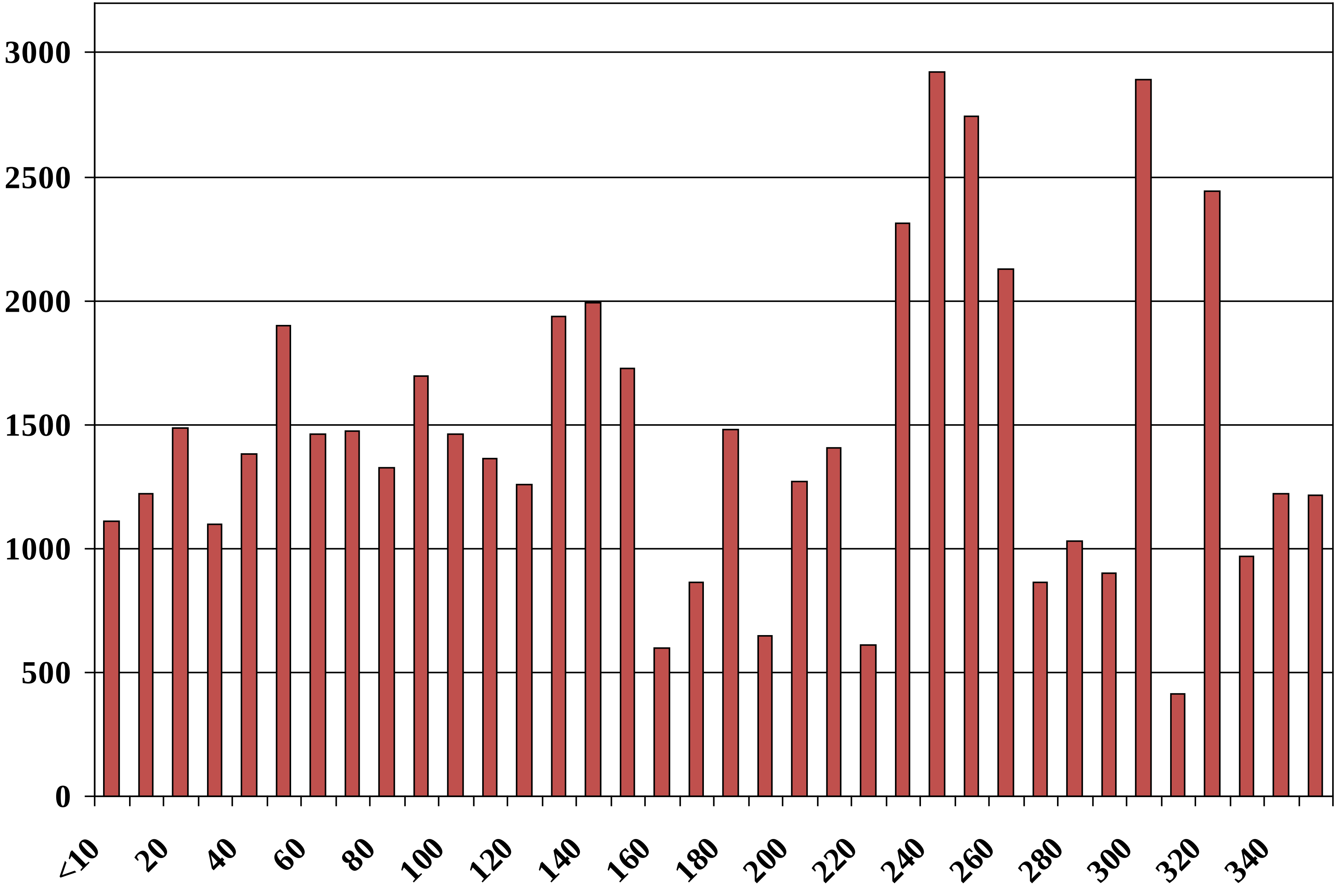


Figure 2. Mean *As* Concentrations (ppb) by Owned Landholdings (square decimeters)



5. Sample for empirical analysis

A. 1,170 respondents aged 18-59 in 465 villages; 583 lineage groups:

B. Tubewells source of drinking water for 97.6% of respondents in 2007/8; in 2002 was 2/3

C. Almost a quarter of households do not use well water for cooking. Avoiding tubewells as a source of cooking water appears to be costly - associated with effort.

The distance to the water source for cooking is a statistically significant 15% higher for users of non-tubewell sources and time spent fetching water in such households is a statistically significant 19.6% higher.

Table 1A
Respondent Characteristics (Means and Standard Deviations):
Men and Women Aged 18-59

Variable	Men	Women
As concentration (ppb)	1367 (1870)	1456 (2225)
Raven's CPM score (number of correct answers)	4.19 (2.09)	3.33 (1.83)
Pinch test pressure (kg)	43.2 (25.9)	31.5 (22.1)
Years of completed schooling	5.26 (4.40)	4.29 (3.96)
BMI	19.5 (2.73)	19.8 (3.10)
Illness in the last week	.189 (.392)	.297 (.457)
Skill occupation	.515 (.500)	.049 (.216)
Annual days worked in the labor market	297.5 (104.8)	15.7 (59.2)
Operate a nonfarm business	.175 (.380)	.010 (.100)
N	742	778

Standard deviation in parentheses.

Table 1B
Food Consumption and Cooking-Water Source, by Gender

Variable	Men	Women
Grain consumption (grams per day)	519.8	448.0
	(244.6)	(177.1)
Green vegetable consumption (grams per day)	37.8	38.7
	(86.5)	(76.2)
Vegetable consumption (grams per day)	137.4	112.5
	(162.8)	(120.5)
Tuber consumption (grams per day)	87.7	76.9
	(88.4)	(76.4)
Fruit consumption (grams per day)	16.8	15.4
	(69.1)	(51.0)
Meat consumption (grams per day)	80.7	61.7
	(99.5)	(74.8)
Number of cigarettes smoked per day	7.38	1.08
	(10.7)	(1.21)
Cooking water source not a well	.232	.242
	(.422)	(.428)
N	742	778

Standard deviation in parentheses.

Estimates of food intake, smoking on A_s retention

A. OLS estimates are rejected statistically (odd results)

B. The IV estimates indicate that consumption of the staple of rural Bangladesh diet (rice) is causally associated with increased retained *As*, conditional on the water source used for cooking and other dietary intakes, and has the largest negative impact of all the consumed food groups:

> *a one-standard deviation increase in grain* (*rice*) *consumption increases arsenic retention by 12.6%.*

C. Given dietary intake, smoking increases arsenic retention, consistent with prior findings in medical literature:

> *the cessation of smoking would lower retained arsenic by 4%.*

D. Three food groups decrease retained arsenic:

> tubers, meats and green vegetables and fruits, consistent with medical findings (e.g., folates enhance methylation)

E. There is a substantial payoff from shifting the source of cooking water from wells:

> *switching from wells to obtain water for cooking evidently decreases retained As by a statistically significant 18.2%.*

F. Men retain over 13% more arsenic than do women, given food intakes, smoking behavior, and common sources of water for cooking and drinking.

Table 2
Individual-Specific Production Function Estimates for (Log) *As* Concentrations, by Estimation Method

Variable/Estimation Method	Village FE	Village FE-IV
Log grain consumption	-.0206	.314
	(0.68)	(3.62)
Log green vegetable consumption	-.0096	-.0309
	(1.71)	(2.33)
Log vegetable consumption	-.0118	-.0197
	(1.64)	(1.39)
Log tuber consumption	-.0117	-.0495
	(0.99)	(1.98)
Log fruit consumption	.0053	-.0085
	(0.64)	(0.50)
Log meat consumption	-.0220	-.0357
	(3.23)	(2.57)
Log number of cigarettes	.0142	.0396
	(1.11)	(2.05)
Cooking water not from a well	-.0907	-.182
	(1.93)	(1.84)
Male	.180	.145
	(3.00)	(2.51)
Male x age	-.0012	-.0030
	(0.64)	(1.63)
N	3,036	3,036
Endogeneity test: *Wu-Hausman* $F(9, 2553)$ [p]	2.99 [.0015]	

Specification also includes the age and age squared of the respondent. Absolute values of *t*-ratios in parentheses.

First-Stage estimates

Does our instrument (mean of the log *As* residuals for non-resident kin) predict well retained arsenic?	Yes
Is the effect non-linear, consistent with the genetics literature (polygenic and epistatic inheritability)?	Yes
Do the effects differ by gender?	No

Figure 3. Locally-weighted Estimates of the Effects of Non-Coresident Lineage *As* Endowments on Respondent *As* Concentrations, by Lineage *As* Endowment Size

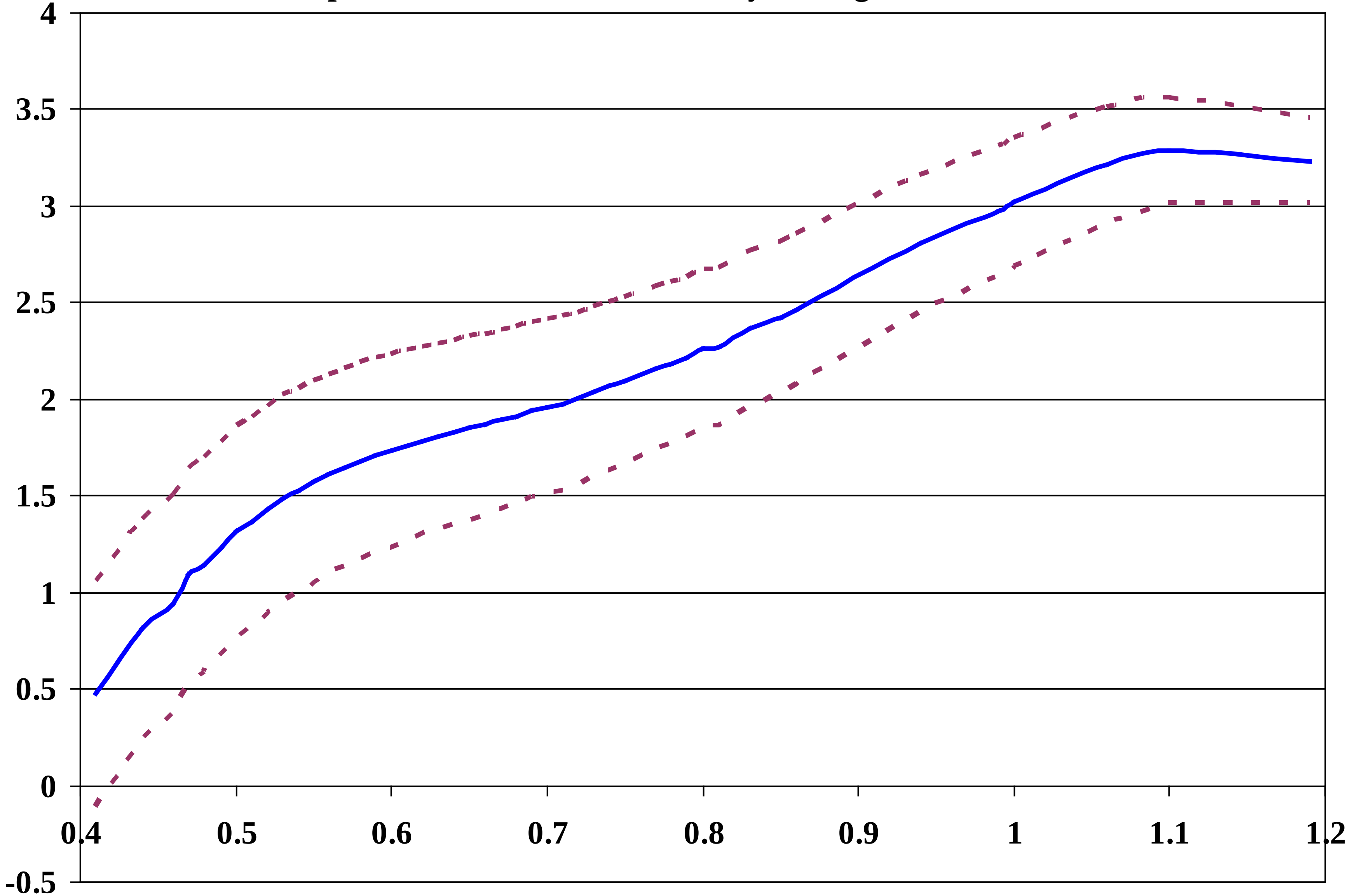


Table 3

First-Stage Coefficient Estimates, Respondents Aged 18-59: Dependent variable = Log *As*

Variable	(1)	(2)
Mean log non co-resident lineage (NCL) *As* residuals	1.86	-6.93
	(2.81)	(1.66)
Mean log non co-resident lineage (NCL) *As* residuals squared	-	5.44
		(2.25)
Age	.0219	.0189
	(1.27)	(1.05)
Age squared	-.0287	-.0250
	(1.19)	(1.01)
Value of owned landholdings ($x10^{-7}$)	.854	.725
	(2.02)	(2.00)
Male	.0291	.0382
	(0.34)	(0.59)
N	1520	1520
F-test endowment instrument coefficients = 0 [*p*]	7.91 [.0054]	6.18 [.0025]
F-test gender coefficients = for all coefficients [*p*]	1.01 [.427]	1.43 [.179]

Absolute values of *t*-ratios in parentheses clustered at the village level.

IV and LIML estimates of δ by gender (use the same first stage estimates regardless of subsample):

Direct productivity measures:	Abridged Raven's CPM test scores, log of pinch strength test results (kg of pressure), schooling attainment (to test validity)
Health:	weight/height, morbidity
Economic outcomes:	occupation, entrepreneurship, labor supply
Economic costs:	Earnings (household expenditires), home productivity
Control variables:	Age, age squared, value of landholdings, average age of co-resident males and females

Findings and diagnostics for Raven's Test performance:

A. OLS estimates (which indicate smaller effects) are rejected.

B. Weak instrument hypothesis rejected.

C. Overidentification test passed, and test has power.

D. Retained arsenic significantly reduces performance for men and women equally:

> *a one standard deviation decrease in arsenic retention would increase performance on the test by one full correct answer, an increase in performance of 24%.*

E. Effects are the same for old and younger respondents: 18-34 and 35-59

Table 4A
Estimates of the Effect of Log *As* on Cognitive Performance: Raven's CPM Score, by Estimation Procedure and Instrument Set for Respondents Aged 18-59

Estimation method	OLS	IV	IV
Instrument set	-	NCL residuals + NCL residuals squared	NCL residuals + NCL residuals sq + NCL *As*
Log *As*	-.173	-.669	-.132
	(3.32)	(6.36)	(1.65)
Age	-.133	-.122	-.134
	(4.76)	(3.91)	(4.86)
Age squared	.109	.0935	.110
	(3.08)	(2.27)	(3.17)
Value of owned landholdings	3.55	3.97	3.51
($x10^{-7}$)	(4.61)	(4.73)	(4.60)
Male	.904	.914	.900
	(8.89)	(7.74)	(8.91)
N	1519	1519	1519

Specification also includes average ages and numbers of male and female household members and number of males aged 18-59. Absolute values of *t*-ratios in parentheses clustered at the village level. NCL=non-coresident lineage.

Table 4B
Test Statistics for Raven's Test Estimates

Estimation method	IV	IV
Instrument set	NCL residuals + NCL residuals squared	NCL residuals + NCL residuals sq + NCL *As*
Endogeneity test: *Wu-Hausman* $F[p]$	17.7 [.000]	0.661 [.412]
Weak identification test: *Cragg-Donald Wald* F	24.8	477.7
Overidentification test: *Hansen J* $\chi^2(1)$ $[p]$	1.05 [.307]	2.70 [.100]
Orthogonality of NCL *As*: *Hansen C* $\chi^2(1)$ $[p]$	-	4.14 [.042]

NCL=non-coresident lineage.

Table 6
LIML Estimates of the Effect of Log *As* on Cognitive Performance by Age Group: Raven's CPM

Age Group		18-34	35-59
Log *As*		-.633	-.755
		(3.73)	(2.68)
Age		.0133	.0189
		(0.08)	(1.20)
Age squared		-.191	-.0251
		(0.60)	(1.16)
Value of owned landholdings (x10^{-7})		4.87	3.17
		(4.39)	(3.74)
Male		1.00	.820
		(7.97)	(6.32)
N		878	641
As gender coefficients =	$\chi^2(1)$ [*p*]	0.22 [.640]	2.86 [.100]
As age-group coefficients =	$\chi^2(1)$ [*p*]	0.19 [.665]	

Specification also includes average ages and numbers of male and female household members and number of males aged 18-59. Absolute values of *t*-ratios in parentheses clustered at the village level.

Are the estimates spurious?

Are genes affecting arsenic methylation correlated with innate intelligence?

A. A cohort test based on schooling attainment

We would expect those with lower performance would obtain less schooling

If performance is not actually affected by arsenic retention, then we would expect the contaminated old and young to both have lower attained (completed) schooling (just a fixed effect)

But many of the older respondents made schooling decisions <u>before</u> the shift to arsenic-contaminated water sources

For the old, current arsenic retention should not affect their schooling; only the young's schooling should be affected

Split the sample by age (23-34 versus 45-64), look at effects of retained *As* on schooling years by age group.

Only the young's schooling is affected; diminished effects by age

Of course, the shift from ground- to well-water occurred gradually over time

We should therefore see the effects of arsenic retention (measured currently) become more negative over time

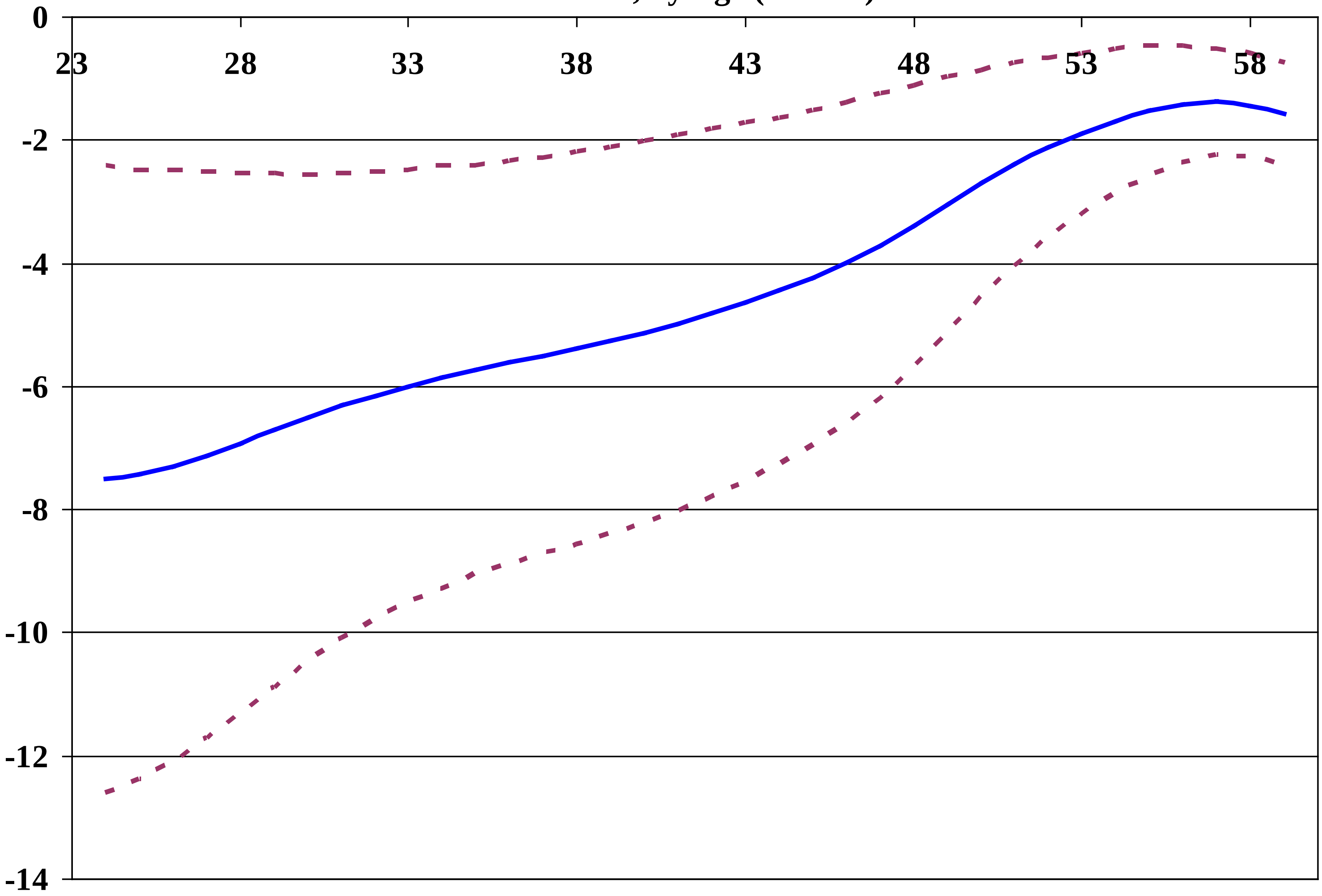
Lowess-smoothed local-IV estimates

Table 7
LIML Estimates of the Effect of Log *As* on Schooling Attainment, by Gender and Age

Gender	Male		Female	
Age group	23-34	45-64	23-34	45-64
Log *As*	-3.57	-.539	-.868	-.226
	(2.16)	(1.25)	(0.66)	(0.51)
Age	-1.88	1.75	.175	-1.59
	(0.95)	(1.57)	(0.09)	(1.72)
Age squared	3.36	-1.67	-.889	1.42
	(0.95)	(1.62)	(0.26)	(1.58)
Value of owned landholdings ($x10^{-7}$)	32.6	6.32	10.9	1.91
	(4.53)	(1.44)	(2.23)	(0.96)
N	248	205	231	188
Endogeneity test: *Wu-Hausman* *F* [*p*]	5.88 [.017]	1.39 [.243]	1.25 [.168]	4.98 [.034]
As coefficients = across age groups within gender $\chi^2(1)$ [*p*]	3.75 [.053]		0.29 [.588]	
As coefficients = by gender within age group $\chi^2(1)$ [*p*]	6.31 [.012]	0.28 [.600]	-	-

Specification also includes average ages and numbers of male and female household members and number of males aged 18-59. Absolute values of *t*-ratios in parentheses clustered at the village level.

Figure 4. Locally-weighted IV Estimates of the Effects of *As* on Years of Schooling Completed for Men, by Age (Cohort)



B. Estimates based on Hapmap (and some elementary molecular biology)

The basic genetic variations examined are single nucleotide polymorphisms (SNPs): DNA sequence variations that occur when a single nucleotide (A (adenine), C (cytosine), T (thymine), or G (guanine)) in the genome sequence is altered.

Two SNPs are said to be in "linkage disequilibrium" (LD) when alleles (the variations) at two or more loci (places on the DNA sequence) appear together more often than would be expected by chance.

LD in humans primarily manifests itself in correlation between pairs of SNPs on the same chromosome and typically extends only for relatively short distances.

The SNPs associated with methylation are close together; but they are not close to SNPs associated with the outcome measures studied.

A review of the literature suggested a set of 18 SNPs for intelligence (five SNPs), body mass and height (eight SNPs), and muscle development and strength (five SNPs).

We use these 18 SNPs in our own LD analysis using human genomic data

The Hapmap data we use:

HapMap Release 22:	Genotypes for more than 3.1 million single SNPs assayed from 30 lineage trios (father-mother-child) of Utah residents with ancestry from Northern and Western Europe.

This subpopulation sample of SNPs are considered relevant for South Asian populations because South Asian Indian populations are both geographically and genetically intermediate between European and East Asian populations.

No correlation (LD) between arsenic SNP's and the 18 SNPs linked in the medical literature to "IQ", strength, body mass.

Does the test have power?

We computed *p*-values for the five pairs of SNPs (three pair of arsenic SNPs and two pair of muscle strength SNPs) that share a gene location and are thus located nearer to each other on the genome: rejection in 3 out of the 5.

Table 8A

Tests of Linkage Disequilibrium between SNPs related to arsenic metabolism and SNPs related to IQ: *z*-statistics for test that r-squared = 0

	Arsenic SNP Number (rs#)						*Holm-Sidak* critical rejection *p*-values
SNP type/ Number (rs#)	11191439	7085104	4925	11509438	1801133	1801131	
IQ							
363039	0.424	0.775	1.225	0.980	0.600	0.949	
4680	1.587	0.245	0.346	0.000[a]	0.346	0.848	
2760118	1.010	1.095	0.000[a]	0.490	2.312	0.916	
821616	0.000[a]	1.428	0.600	1.225	1.625	1.342	
1018381	1.249	1.929	0.648	0.490	0.346	0.648	0.531
All SNPs							0.487

[a] r-squared rounded to 0.000 by PLINK program. Data from the International HapMap project database, CEU founders (release 22) (http://hapmap.ncbi.nlm.nih.gov/index.html.en). Computed with PLINK software from the Broad Institute.

Table 8B
Tests of Linkage Disequilibrium between SNPs related to arsenic metabolism and SNPs related to Body Mass and Height: *z*-statistics for test that r-squared = 0

	Arsenic SNP Number (rs#)						*Holm-Sidak* critical rejection *p*-values
SNP type/ Number (rs#)	11191439	7085104	4925	11509438	1801133	1801131	
Body mass and height							
724016	0.000[a]	0.346	0.774	0.648	0.916	1.068	
143384	0.980	2.064	0.812	0.000[a]	1.296	2.738	
1351394	0.245	0.548	1.200	1.661	0.774	0.000[a]	
7689420	0.000[a]	0.000[a]	0.600	1.225	1.200	1.095	
6449353	0.245	0.245	1.990	0.490	1.897	0.245	
1421085	0.490	0.693	1.095	1.470	2.349	0.346	
211683	0.735	0.245	0.245	1.296	1.849	2.683	
988712	0.245	0.648	0.774	1.010	1.549	0.848	0.229
All SNPs							0.487

[a] r-squared rounded to 0.000 by PLINK program. Data from the International HapMap project database, CEU founders (release 22). Computed with PLINK software from the Broad Institute.

Table 8C

Tests of Linkage Disequilibrium between SNPs related to arsenic metabolism and SNPs related to Physical Strength: *z*-statistics for test that r-squared = 0

	Arsenic SNP Number (rs#)						*Holm-Sidak* critical rejection *p*-values
SNP type/ Number (rs#)	11191439	7085104	4925	11509438	1801133	1801131	
Muscle development and physical strength							
7843014	0.490	0.000[a]	0.424	1.200	0.245	0.548	
7460	0.735	0.346	0.693	0.774	1.296	0.000[a]	
1800169	0.916	0.245	1.587	0.693	1.897	0.000[a]	
1815739	0.693	0.648	1.661	0.346	0.245	1.944	0.798
3808871	0.000[a]	0.548	0.735	0.245	1.068	0.346	
All SNPs							0.487

[a] r-squared rounded to 0.000 by PLINK program. Data from the International HapMap project database, CEU founders (release 22). Computed with PLINK software from the Broad Institute.

Findings for the effects of *As* retention on strength:

A. OLS underestimates the effects of arsenic retention, just as for cognition.

B. Significant reduction that is statistically significant for men, but cannot reject the same for women.

a one-standard deviation increase in retained arsenic reduces physical performance by over 6%

Findings for self-reported, coventional morbidity symptoms (e.g., headaches, diarrheal symptoms, fever, and coughing) and body mass:

<u>No</u> effects or arsenic retention on these standard measures of health

The effects of arsenic retention are physically “invisible”

Table 5
LIML Estimates of the Effect of Log *As* on Physical Performance (Pinch Test) for Respondents Aged 18-59, by Gender: *Kg* of pressure

Gender	Both	Male	Female
Log *As*	-1.60	-3.14	-1.05
	(1.97)	(1.97)	(0.56)
Age	.662	.611	.923
	(2.69)	(2.73)	(2.37)
Age squared	-.991	-.854	-1.39
	(2.56)	(2.34)	(2.50)
Value of owned landholdings ($x10^{-7}$)	.776	1.09	2.53
	(0.23)	(0.71)	(0.62)
Male	11.7	-	-
	(8.77)		
N	1519	777	742
As gender coefficients = $\chi^2(1)$ [*p*]		0.51 [.716]	

Specification also includes average ages and numbers of male and female household members and number of males aged 18-59. Absolute values of *t*-ratios in parentheses clustered at the village level.

Table 9
Estimates of the Effect of Log *As* on Morbidity and Log Body Mass for Respondents Aged 18-59, by Estimation Method

Gender	Illness in the Last Week		Log BMI	
Estimation method	Probit	LIML Probit	OLS	LIML
Log *As*	-.0353	.0902	-.0139	-.0168
	(1.09)	(0.91)	(2.16)	(0.50)
Age	-.0356	-.0380	.0161	.0162
	(1.92)	(2.05)	(6.50)	(7.08)
Age squared	.0675	.0706	-.0213	-.0214
	(2.87)	(3.01)	(4.42)	(6.52)
Value of owned landholdings ($x10^{-7}$)	-.181	-.287	.291	.294
	(0.46)	(0.71)	(4.42)	(4.15)
Male	-.370	-.370	-.0147	-.0146
	(5.99)	(6.05)	(2.20)	(2.16)
N	1519	1519	1519	1519
Endogeneity test: *Wald* $\chi^2(1)$ [*p*]	2.09 [.148]		-	
Endogeneity test: *Wu-Hausman F* [*p*]	-		0.0087 [.921]	

Specification also includes average ages and numbers of male and female household members and number of males aged 17-59. Absolute values of *t*-ratios in parentheses clustered at the village level.

Economic outcomes

Given the reduction in cognitive performance and schooling for the young, would expect to see adverse effects in economic performance.

LIML findings for <u>individual</u> occupation, entrepreneurship (operate a nonfarm business), work time for men aged 23-34:

A. <u>Skill occupation</u>: where decision-making, thinking important (teacher, doctor, government administrator) or in business management including farmers (but excluding a farm worker) and shopkeepers (49%):

cutting by half the average levels of arsenic in this sub- population would increase the proportion of men in skilled occupations by 12.2 percentage points, or by 24%.

B. Entrepreneurship (19.2%):

cutting retained arsenic levels by the same 50% would increase the proportion of men running nonfarm businesses by over 5 percentage points, an increase of 26%.

C. Labor supply (total annual days worked):

No effect on time worked.

But is there a productivity effect?

Table 10
Estimates of the Effect of Log *As* on Occupational Choice, Entrepreneurship and Labor Supply: Males Aged 23-34, by Estimation Method

Dependent variable:	Skilled Occupation		Operate a Nonfarm Business		Total Annual Days Worked in the Labor Market	
Estimation method	Probit	LIML Probit	Probit	LIML Probit	OLS	LIML
Log *As*	.00244	-.628	-.149	-.971	-17.8	6.34
	(0.02)	(5.30)	(1.55)	(11.2)	(2.07)	(0.55)
Age	1.13	.873	.804	.344	3.83	4.07
	(1.84)	(1.81)	(1.20)	(0.84)	(0.12)	(0.13)
Age squared	-1.98	-1.52	-1.35	-.573	-1.92	-2.84
	(1.86)	(1.81)	(1.17)	(0.82)	(0.03)	(0.05)
Value of owned landholdings	5.20	4.99	-2.80	-.395	76.5	48.0
($x10^{-7}$)	(2.76)	(2.60)	(1.59)	(0.35)	(0.81)	(0.56)
N	260	260	260	260	260	260
Endogeneity test: *Wald* $\chi^2(1)$ [*p*]	2.14 [.144]		2.09 [.148]		-	
Endogeneity test: *Wu-Hausman F* [*p*]	-		-		5.51 [.021]	

Specification also includes average ages and numbers of male and female household members and number of males aged 18-59. Absolute values of village-clustered *t*-ratios in parentheses.

Estimating *As* effects on earnings in a household context using genetic linkages

The household expenditure equation that we estimate is

(26) $$\text{Log}E_j = \beta_1 A_{ijm} + \beta_2 A_{ijm} N_m + \beta_3 A_{ijf} + \beta_4 A_{ijf} N_f + \boldsymbol{Z\beta}_5 + \varepsilon_j,$$

where $A_{ijm(f)}$ =log arsenic retention of prime-age men (women),

$N_{m(f)}$=number of prime-age men (women),

$\boldsymbol{Z}$ is a vector of control variables: total number, age and sex composition of household members.

Given the division of labor in Bangladesh households, we would expect that β_3, β_4=0, as women do not participate in the labor force.

The model indicates that $\beta_2 = w_j \Omega h_1 < 0$, which is the earnings loss for a male earner from an increase in retained arsenic (available for all households with male earners).

Cannot estimate (26) using OLS.

Identification issues from using kin-based instrument:

> If all members of the household (men and women) are in the same lineage, cannot identify or separate β_2 and β_4 - same instrument for men and women.
>
> The effect of variation in retained arsenic within gender and age groups is the same for all members of the lineage in the household.

Two conditions are necessary:

1. *Exclusivity* condition: there are some households that only have members within the same lineage and gender/age group and not members of the same lineage in other age/gender groups.

Easily met in the data:

Few prime-age sisters or daughters of heads co-reside. This enables identification of β_1 and β_2 (families with prime-age males in same lineage, no women).

Few wives co-reside with their father or brothers, and not all have prime age adult sons, β_3 and β_4 are identified (families with no male kin of wives).

2. *Orthogonality* condition: problem that we do not have lineage instruments for all members

Lineage instrument for head, but not wife (since she may have not been in prior round of survey), but if uncorrelated then no bias.

Thus, assume no assortative mating on arsenic genes (hidden).

Cohort test of assortative mating on individual genetic ability to methylate:

> Those couples who married when arsenic-contaminated water was not a health problem, prior to the 1980's, clearly did not sort on propensities to methylate arsenic.
>
> If such sorting did occur, it would have been after the problem was well known.
>
> We would then expect that the current relationship between the retained arsenic of husbands and wives would be stronger among couples who married in recent years compared with couples who married before the 1980's.

Table 11

Within-village Relationship between Husbands and Wives Log *As*, by Period When Married: All Marriages Occurring Before 1981 and After 1990

Variable	(1)	(2)
Log husband's *As*	.407	.402
	(5.73)	(6.02)
Log Husband's *As* x married before 1981	-	.0149
		(0.32)
Married before 1981	-	-.00716
		(0.02)
N	1114	1114

Absolute values of *t*-ratios in parentheses clustered at the village level. Village fixed effects included in the specification. Pre-1981 correlation = .840; post-1990 correlation=.776.

We estimate the household expenditure equation(26) for 720 households, which are required to have at least one prime-age adult, aged 18-59, of either gender.

Of these, 459 (64%) meet the exclusivity criterion and contribute to identification of the β's - there are men or women in the relevant age group with different lineages

Findings:

A. The test statistics indicate only the retained arsenic of prime age males matter for total household expenditure, consistent with the division of labor.

B. Given that time worked appears to be insensitive to changes in retained arsenic (Table 10), the β_2 point estimate, which is estimated precisely, indicates that

reducing arsenic levels to those in the United States would increase male market productivity (annual earnings) in rural Bangladesh by 9%.

Table 13A
Estimates of the Effect of *As* Contamination within the Household on Log Total Annual Household Expenditures, by Estimation Method

Estimation Method	OLS	LIML
Mean Log *As* of Men Aged 18-59 (m*As*)	-.107	.00113
	(0.85)	(0.17)
Mean Log *As* of Men Aged 18-59 x Number of Men Aged 18-59 (m*As* x mp)	.0279	-.0941
	(1.55)	(2.65)
Mean Log *As* of Women Aged 18-59 (f*As*)	.0481	.00109
	(1.48)	(0.14)
Mean Log *As* of Women Aged 18-59 x Number of Women Aged 18-59 (f*As* x fp)	-.0300	-.00108
	(1.65)	(0.05)
Number of Men Aged 18-59 (mp)	-.107	.673
	(0.85)	(2.77)
Number of Women Aged 18-59 (fp)	.233	-.00393
	(1.75)	(0.03)
Value of owned landholdings ($x10^{-7}$)	1.54	1.91
	(6.82)	(9.73)
N	720	720

Specification also includes average ages and numbers of male and female household members. Absolute values of *t*-ratios in parentheses clustered at the village level.

Table 13B
Test Statistics: Estimates of the Effect of *As* Contamination within the Household on Log Total Annual Household Expenditures

Test		Test Statistic
ρ m*As* and expenditure residuals		.222
		(1.46)
ρ (m*As* x mp) and expenditure residuals		.262
		(1.63)
ρ f*As* and expenditure residuals		.173
		(1.21)
ρ (f*As* x fp) and expenditure residuals		.169
		(1.15)
Test: error correlations = 0 (endogeneity of *As*)	$\chi^2(4)$ [*p*]	7.15 [.128]
Test m*As* = 0, m*As* x mp = 0	$\chi^2(2)$ [*p*]	7.46 [.024]
Test f*As* = 0, f*As* x fp = 0	$\chi^2(2)$ [*p*]	0.02 [.990]
Test f*As* x fp < m*As* x mp =	$\chi^2(1)$ [*p*]	4.98 [.013]

Absolute values of *t*-ratios in parentheses clustered at the village level.

Is there an effect of arsenic retention among women on household productivity?

We now test propositions 2 and 3 of the model by estimating the determinants of whether or not the household chooses a non tubewell source of cooking water using the same sample of households.

The equation we estimate is

$$(27) \qquad W_j = \gamma_1 A_{ijm} + \gamma_2 A_{ijm} N_m + \gamma_3 A_{ijf} + \gamma_4 A_{ijf} N_f + \mathbf{Z}\,\gamma_5 + \varsigma_j,$$

where W_j takes on the value of 1 if the household does not use tubewell water for cooking.

Recall that non tubewell sources of water are more costly, requiring more time than tubewell sources for water consumption, but that use of such sources significantly reduce retained arsenic (Table 3).

Derived from the model:

Proposition 2: if higher arsenic retention among women is associated with increased use of nontubewell sources of water this would imply that retained arsenic among women reduced productivity in producing home goods; i.e., $\gamma_4>0$, given that only women fetch and treat water.

Proposition 3: the sign of $\gamma_{2,}$ the effect of male-earner reatined arsenic on water source choice, indicates whether household and purchased goods were substitutes ($\gamma_2<0$) or complements ($\gamma_2>0$), given that no men contribute significantly to household production.

Findings from the choice of cooking water source:

A. The OLS estimates of the γ's are negatively biased - reflect our finding that nontubewell water sources decrease arsenic (reverse causation).

B. The LIML γ estimates indicate that arsenic contamination significantly reduces productivity in household goods production and that household goods and purchased goods are *complements.*

if the reduction in home and market productivity are similar, the 9% reduction in household expenditures corresponds to the overall reduction in household consumption, and not just that part which is transacted in the market.

Table 14A
Estimates of the Effect of *As* Contamination within the Household on Whether the Household Chooses non Tubewell Water, by Estimation Method

Estimation Method	Probit	LIML Probit
Mean Log *As* of Men Aged 18-59 (m*As*)	-.0230	.0030
	(0.81)	(0.11)
Mean Log *As* of Men Aged 18-59 x Number of Men Aged 18-59 (m*As* x mp)	.136	.399
	(1.30)	(4.87)
Mean Log *As* of Women Aged 18-59 (f*As*)	.0777	.0883
	(2.99)	(3.24)
Mean Log *As* of Women Aged 18-59 x Number of Women Aged 18-59 (f*As* x fp)	.238	.362
	(2.80)	(4.50)
Number of Men Aged 18-59 (mp)	-.987	-2.73
	(1.51)	(5.44)
Number of Women Aged 18-59 (fp)	-1.93	-2.68
	(3.22)	(5.18)
Value of owned landholdings ($x10^{-7}$)	-.319	-1.12
	(0.50)	(1.81)
N	1101	1101

Specification also includes average ages and numbers of male and female household members. Absolute values of *t*-ratios in parentheses clustered at the village level.

Table 14B

Test Statistics: Estimates of the Effect of *As* Contamination within the Household on Whether the Household Chooses non Tubewell Water

Test		Test Statistic
ρ m*As* and cooking water residuals		-.654
		(5.63)
ρ (m*As* x mp) and cooking water residuals		-.713
		(6.38)
ρ f*As* and cooking water residuals		-.634
		(5.31)
ρ (f*As* x fp) and cooking water residuals		-.679
		(5.85)
Test: error correlations = 0 (endogeneity of *As*)	$\chi^2(4)$ [*p*]	16.9 [.002]
Test: m*As* = 0, m*As* x mp = 0	$\chi^2(2)$ [*p*]	34.8 [.000]
Test: f*As* = 0, f*As* x fp = 0	$\chi^2(2)$ [*p*]	52.1 [.000]
Test: f*As* x fp < m*As* x mp =	$\chi^2(1)$ [*p*]	0.08 [.389]

Absolute values of *t*-ratios in parentheses clustered at the village level.

Policy implications: costs and benefits

A. <u>Changes in diet</u>:

While we have found that diet matters for arsenic retention, *the beneficial effects of dietary changes are relatively small*

A doubling of tuber consumption would only reduce retained arsenic by 5%.

Cessation of smoking for smokers: *less than 4% reduction in As*

B. Switching sources of cooking water:

If everyone switched from tubewells as their source of cooking water, retained arsenic would decline by 18%,

But, that would only represent less than 20% of the gap between average retained arsenic levels in the rural Bangladesh population and that in non-contaminated populations

would only increase incomes by less than 2%.

C. Eliminating arsenic from all water consumed: what level of investment per person is justified by the benefits (e.g., piped water from clean source)?

Annual Benefits:

Male workers:	$54
Women: if ½ as productive in home but same percentage loss:	$27
a reduction in women's time spent fetching water (20 minutes per day in our data on average).	$13

PDV over 20 years, 3% to 8% discount rate: *$1000-$1400*

Costs of no action increase if there is economic growth