Identifying the Hidden Costs of a Public Health Success:
Arsenic Well Water Contamination and Productivity in Bangladesh

Mark M. Pitt
Brown University

Mark R. Rosenzweig
Yale University

Nazmul Hassan
University of Dhaka

June 2012

The research reported in this paper was supported by grants from the National Institutes of Health, ROlDK072413, the International Growth Center, RA-2009-ll-028, the William and Flora Hewlett Foundation, 5-29193, and from the MacMillan Center, Yale University. We are grateful to Katherine Kelley of the University of Rhode Island and Dave Murray and Bob Correra of Brown University for their excellent lab work, to Eva Kolker for research assistance and to Pascaline Dupas for insightful comments on an earlier draft.

1. Introduction

In the 1970s and 1980s, the government of Bangladesh, with the support and financing of the United Nations Children's Fund, promoted the digging of tube wells to provide clean drinking water and reduce the incidence of diarrheal disease. Prior to this, drinking water came from surface water sources, which were identified as a principal source of diarrheal disease. This shift in water sources has been attributed to having played a major role in the reductions in morbidity experienced in Bangladesh, particularly with respect to water-borne diseases. Indeed, despite little growth in rural wages and no change in average calorie intake (Pitt *et al.*, forthcoming), there has been significant increases in average BMI by age and in height for the rural population of Bangladesh, as shown in Appendix Figures A and B for the period 1982 and 2008 from panel survey data based on a representative population of rural households in 1982.

In the late 1990s, however, evidence indicated that groundwater, unlike surface water and by then the main source of water for drinking, irrigation and cooking in Bangladesh, was contaminated by naturally-occurring arsenic in 59 of the country's 64 districts. The contamination of groundwater by arsenic in Bangladesh is the largest poisoning of a population in history (Smith *et al.*, 2000). As 95% of the country's population relies on well water, an estimated 57 million people have been chronically exposed to drinking water with arsenic levels exceeding the WHO standard of 10 μg/L, and 35 million people were exposed to arsenic levels above the country's government standard of 50 μg/L (British Geological Survey, 1999).

While there is now a large body of evidence that the consumption of arsenic-contaminated water is manifested in new diseases and symptoms, mostly related to skin lesions, there is little evidence on the economic consequences of arsenic ingestion and retention in the body in Bangladesh. Unlike for diseases such as dysentery and diarrhea, the debilitating effects of arsenic on productivity are not conspicuous, are largely hidden, if they exist. Such effects are also difficult to estimate, given that arsenic levels in humans reflect consumption and time-allocation choices and much of the population does not work for wages. In this paper we exploit recent molecular genetics evidence on the genetic basis of arsenic excretion and unique information on family links among respondents living in different environments from a large panel survey within a theoretical framework incorporating optimizing behavior to uncover the hidden costs of arsenic poisoning in Bangladesh. In particular, we provide new estimates of (i) the effects of the consumption of foods grown and cooked in arsenic-contaminated water on individual arsenic concentrations and of (ii) the effects of the ingestion and retention of inorganic arsenic on direct measures of cognitive and

physical capabilities as well as on the schooling attainment, occupational structure, entrepreneurship and incomes of the rural Bangladesh population.

The estimates are based on arsenic biomarkers obtained from a sample of members of rural households in Bangladesh who are participants in a long-term panel survey following respondents and their coresident household members over a period of 26 years. In the last round of the survey each respondent in the survey aged eight years and above provided clippings from all ten toenails. Toenail clippings are the preferred biomarkers for measurements of arsenic ingestion in environments in which exposure lasts more than a few months and where a significant share of arsenic exposure occurs through food consumption. Our new estimates of the causal impacts of arsenic exploit the new biomedical evidence on genetic linkages among kin in the ability to methylate arsenic, thus reducing its toxic effects, and data identifying family linkages among respondents living in separate environments (villages).

The link between arsenic exposure and an individual's health and productivity has two components: ingestion and retention. Arsenic is ingested by drinking water and by consuming foods grown and cooked in arsenic-contaminated waters. Estimates of the relative importance of the two sources, based on the distribution of arsenic among wells in Bangladesh and the non-linear relationship between arsenic ingestion and exposure levels, suggest that dietary sources of arsenic are the most important route of arsenic exposure for two-thirds of the Bangladesh population (Kile *et al.* 2007). The toxicity and retention of arsenic that is ingested, however, is also affected by the nutrients contained in the food consumed. Specific foods differ in the degree to which they retain arsenic used in their cultivation and cooking and differ as well in the extent to which, via their nutritional composition, they enable the metabolism (detoxification) and subsequent excretion of arsenic, primarily through a process of methylation. For example, Zablotska *et al.* (2008) and Heck *et al.* (2007), based on randomized field experiments, highlight the protective effects of folate-related nutrients and B-vitamins in samples from rural Bangladesh. Leafy vegetables are known to contain high concentration of folic acid. However, Hossain (2006), based on extensive tests in Bangladesh, reports that among the food products tested, leafy vegetables have the highest concentration of arsenic. The net effect of the consumption of different foods on arsenic retained in the body is thus not known.

A large number of studies have related arsenic exposure in the environment, measured by well contamination, and diet to biomarkers for arsenic in humans (urine, blood, hair, and toenails)

and to specific health outcomes. However, these studies have a number of deficiencies. Neither proximity to arsenic sources (wells) nor the ingestion of arsenic can be considered exogenous to the outcomes studied. First, especially given campaigns to identify wells that are contaminated, households can choose to avoid arsenic contamination by choosing surface-water sources that are less convenient, perhaps then risking the incidence of water-born disease (Fields *et al.*, 2010). Second, arsenic ingestion is related to the nutrient composition of foods, which also reflect preferences and income constraints. An individual's retained arsenic in a country such as Bangladesh, where millions of wells have been tested and identified by their levels of arsenic contamination and where diet is both a source of arsenic and an aid to its metabolism, is clearly not an exogenous variable. Finally, studies that carry out randomized nutrient supplementation that can reduce arsenic's burden by increasing the rate at which it is metabolized (methylated) do not identify the mechanisms by which such arsenic-reducing interventions improve health. Such supplements improve health and thus productivity by more than just reducing the toxic concentration of arsenic and its metabolites in body tissue.

There is little evidence on the effects of arsenic contamination on performance measures, and none on economic outcomes.[1] In recent years, however, the specific set of inheritable genes associated with the abilities of humans to metabolize and secret ingested arsenic have been identified, with field studies showing that retained arsenic levels are correlated within families consistent with genetic models. Genetic variation in abilities to methylate arsenic is thus an important component of the large variance in arsenic found in humans. But this linkage has not been exploited in studies examining the consequences of arsenic. Knowledge of both how arsenic retention directly and causally affects health and productivity outcomes and how diet affects arsenic retention is obviously critical to any cost-benefit analyses of governmental programs aimed at improving health and productivity. But identification of how arsenic ingestion causally affects health and productivity must take into account all of the fundamental determinants of its toxicity, which include behavioral, environmental and genetic components as well as the distribution of arsenic concentrations in the population.

[1]Asadullaha *et al.* (2011) find that proximity to contaminated wells is associated with lower test scores among school-age children, but identification rests on the assumption that proximity to water sources is random, and does not take into account that households may select their water sources endogenously.

In section 2 of the paper we set out a simple model that highlights the distinct endogenous and exogenous components of an individual's retained arsenic. We show that under realistic assumptions about the information set of agents, including that productivity effects of retained arsenic are hidden but health effects are common knowledge, it is not possible to sign the bias in correlational relationships between retained arsenic and productivity outcomes due to optimizing with respect to water choice. We then go on to build a model that incorporates realistic features of the rural Bangladesh household that shows how it is possible to identify the causal effects arsenic ingestion on earnings using household expenditure data. This is important in Bangladesh, as less than half of the male labor force are wage earners. Given that we have consumption expenditures for all households, we can avoid potential selectivity problems that plague estimates of earnings effects from improving health based solely on the earnings of wage workers.

Another feature of Bangladesh households is that almost no women are earners, but women play a large role in household production. We show, using the model, that it is possible identify if arsenic also reduces home productivity and whether home goods are complements or substitutes with purchased goods by estimating the effects of the retained arsenic of men and women on households' choices of water sources. We also show that knowledge of these parameters is necessary to make inferences about the potential welfare effects of arsenic reduction based on earnings estimates as well as inferences about the consequences of efforts to provide cleaner water sources.

Section 3 describes the data and our measure of arsenic retention from the sample of toenail clippings. The data indicate very high levels of arsenic concentrations, exceeding average levels measured in US respondents by almost 20 to one, with concentrations exhibiting wide individual variation but spread almost uniformly across landholding groups. In section 4 we discuss how we use information on the measured retained arsenic family members residing in different villages and estimates of the effects of food consumption on arsenic retention and ingestion to identify the causal effects of arsenic retention on a variety of outcomes at the individual and household level.

To implement the identification method, we first obtain estimates of the causal effects of the individual consumption of seven food groups on arsenic concentrations based on individual-specific food consumption information, water source choice and smoking on the toenail-based arsenic measures. These show that diet matters for arsenic contamination, with the consumption of grains, the staple of the Bangladesh diet, significantly increasing, but tubers, in accord with the medical

literature, reducing measured arsenic concentrations. We also find that using non tubewell water sources for cooking reduces arsenic concentrations by 18%.

In the next sections we examine the relationship between retained arsenic and measures of individual capabilities, schooling, health and economic choices using as instruments measures of arsenic retention net of the influence of diet, water-source choices, and village fixed effects of separated family members. We find that OLS estimates significantly understate the negative effects of retained arsenic on cognition and physical strength. We also show that the negative effects on cognition are manifested in lower schooling attainment especially for males, but only for those cohorts of males who were of school age after the switch to tubewells as water sources and the discovery of arsenic contamination in Bangladesh. Young men with higher concentrations of retained arsenic were also significantly less likely to have a skill occupation where decision-making is important or to be entrepreneurs. We show that these results are not due to a direct correlation in genes associated with arsenic methylation and genes associated with cognition and strength by estimating correlations among the relevant genes using genomic data. Using the same methods, we could find no relationship, however, between BMI or conventional morbidity symptoms, consistent with the productivity effects of arsenic contamination being hidden in the population.

In the penultimate section, we implement the instrumental-variables method in a multi-member household context that is consistent with the household model we constructed in section 2. We show that our data meet the conditions for identifying the causal effects of retained arsenic by gender on earnings using household expenditure data. Our estimates indicate that lowering the amount of retained arsenic among Bangladesh prime-age males to those levels in uncontaminated countries would increase earnings by 9%. We find that retained arsenic also reduces the productivity of women in home production, and that home goods and purchased goods are complements. These latter results imply that the estimated increase in total household expenditures through eliminating arsenic closely corresponds to the overall consumption benefits to the household. They also imply, however, that improvements in water quality at the source would be less effective than measures that reduce the retention of ingested arsenic due to the endogenous response of households in their own efforts to reduce arsenic ingestion. In the conclusion we summarize our results and attempt to quantify the productivity benefits from the provision of arsenic-free water to Bangladeshi households.

2. Modeling the Effects of Arsenic Contamination

a. *Individual Effects of Arsenic Retention*

In this section we describe a simple one-person model to fix ideas about how heterogeneity in water quality, preferences, health information, and abilities to methylate ingested arsenic combined with optimizing behavior affect inferences about the relationship between an individual's measured amount of retained arsenic[2] and measures of capabilities and economic outcomes. In particular, in accordance with the medical literature, retained arsenic A_{ij} for individual i residing in environment j depends on the consumption of specific foods, where $\boldsymbol{C_{ijx}}$ is a vector of consumed foods indexed by x; on the quality (arsenic content) of the water used for cooking and drinking ω_{ij}; and on the individual's endowed ability to methylate arsenic, given by μ_{ij}.

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

We assume that $A_2<0$; $A_3<0$; and A_{22}, A_{33}, $A_{32}>0$, so that higher ω_{ij} indicates less contamination by arsenic.[3] 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 $\omega_1>0$, $\omega_{11}<0$. 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, such as surface water sources, and/or time spent collecting additional fuel needed for water boiling.[4]

The budget constraint is given by (3)

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

where F_{ij}=non-earnings sources of income and the p_{jx} are local food prices. The individual's wage rate w_{ij} depends on the individual's capability or skill h_{ij}, which has a local per-unit rental price w_j, so

[2]Retained arsenic in this context is the concentration of arsenic in body tissue. The measure used in the empirical work is the arsenic concentration of toenails, the preferred biomarker of arsenic retained in human tissue.

[3]The amount of retained arsenic depends on the quantity of water consumed. For simplicity we assume that this is constant across individuals of given age and gender. Note that A_1 cannot be signed *a priori* as foods are both a source of arsenic and, through the nutrients they contain, an aid to its metabolism.

[4]In our survey data, 26.4% of households obtain their water for cooking from sources other than tube wells. Of these households, over three fourths treat the water, the majority by boiling and the rest by an additive such as alum. As discussed below households using non-tubewell-sourced cooking water spend significantly more time fetching water and fuel.

that $w_{ij} = w_j h_{ij}$, and skill is affected negatively by retained arsenic and positively by a skill endowment r_{ij}:

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

where $h_1<0$.

In this set-up, the sources of exogenous variability in retained arsenic are e_j, which is locale-specific, and μ_{ij}, which is person-specific. Given these relationships, 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_j) j is:

(5) $$dh_{ij}/dA_{ij} = h_A(1 + A_2\omega_1(dt_{ij}/d\mu_{ij})(d\mu_{ij}/dA_{ij}) + A_1(dC_{ij}/dF_{ij})(dF_{ij}/dh_{ij})),$$

which reflects the association between observed individual arsenic and the exogenous and endogenous components of arsenic retention. If variation in arsenic retention affects either the composition of goods consumed or water purification effort the association between retained arsenic and measures of capabilities will not identify the causal effect of arsenic retention. Similarly, the association between measures of the quality of water consumed and A_{ij} also reflects behavior, and not just the technological relationship given by (1):

(6) $$dA_{ij}/d\omega_{ij} = A_2 + h_A w_j \omega_1 dt_{ij}/dw_{ij}$$

The biases in the associations (5) and (6) - the terms associated with the behavioral response to the exogenous components of arsenic *(As)* ingestion and retention - depend on how retained arsenic affects behavior. But is it plausible that arsenic affects behavior if the costs and even the amounts of arsenic ingested or retained are unobserved by agents and agents do not even know how or whether that arsenic affects their productivity? We now show that under plausible information assumptions changes in arsenic retention or in available water quality will affect water purification effort. In particular, we 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 and (iii) they know how water purification effort affects arsenic retention. However, they are *un*informed about both (i) the effects of *A*s on productivity *h*, (ii) their own individual arsenic retention A_{ij} and (iii) their own methylation efficiency endowment μ_{ij}. They only observe their own productivity. The individual effects of arsenic retention are hidden.

For simplicity we ignore for now the effect of the consumption of specific foods on arsenic ingestion and consider the consumption of only one aggregate consumption good C_{ij} with a unit

market price. Changes in effort are thus the only source of endogeneity in this case. The utility function is

(7) $$U = U(A_{ij}^*, C_{ij}; u_{ij}) \qquad U_1<0, U_2>0, U_{11}>0, U_{22}<0,$$

where A_{ij}^* = the agent's beliefs about his retained arsenic and we have added an individual-specific preference parameter u_{ij}.

What is the optimal time spent fetching water, given the beliefs of agents? The FOC is:

(8) $$U_A A_2^* \omega_1 = U_C w_{ij},$$

Agents face a trade-off between consumption (good) and arsenic (bad) because effort reduces the time available for income earning. Note that if agents were informed about the economic consequences of arsenic retention the FOC is:

(9) $$U_A A_2^* \omega_1 = U_C(w_{ij} - A_2^* \omega_1 h_1 w_j(\Omega - t_{ij}));$$

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 effort) than is optimal, if lowering arsenic retention actually increases productivity, as $U_C w_{ij} > U_C(w_{ij} - A_2^* \omega_1 h_1 w_j(\Omega - t_{ij}))$. There is thus a productivity and health payoff to the dissemination of information about the economic consequences of arsenic retention, if there are such consequences.

In this simple model, given realistic information constraints, can the bias in (5) and (6), which depends on how arsenic affects water purification effort, be signed?

Proposition 1:

The effect of retained arsenic on the demand for improved water via increased purification effort is ambiguous. The bias in the association between individual arsenic retention and productivity thus cannot be signed.

Proof: Assume for simplicity that A and C are separable in U, then

(10) $$dt_{ij}/d\mu_{ij} = h_1 A_3[U_C w_j/\Phi + dt_{ij}/dF_j w_{ij}(\Omega - t_{ij})],$$

where $\Phi = -[(A_2\omega_1)^2 U_{AA} + U_A A_{22}(\omega_1)^2 + A_2\omega_{11}] < 0$ and $dt_{ij}/dF_j = U_{CC}/\Phi > 0$

The first term in (10) is negative - lower arsenic retention increases market productivity and the opportunity cost of effort. The second term is the income effect, which is positive on purification effort - the increase in income (or C) lowers the marginal utility of consumption relative to that of arsenic.

Improvements of the local water source (change in e_j) also affect behavior, except that in principle, agents can be told the level of e_j so that there is an additional negative public-health

information effect on efforts to reduce arsenic ingestion, which is embodied in the first term in brackets in (11):

$$(11) \qquad dt_{ij}/de_j = \omega_1[(A_2)^2 U_{AA} + U_A A_{22}]/\Phi + A_2 h_1[U_C w_j/\Phi + (\Omega - t_{ij}) w_j dt_{ij}/dF_j]$$

Given (11), the effect of an intervention that reduced arsenic in the local water source (by, say, drilling and testing a new tubewell) on income ($dF_j/de_j = dC_j/de_j$) may thus under or over-estimate the pure productivity effect, and the total welfare gain, of such an intervention, which is $(\Omega - t_{ij})h_1 A_2 w_j$. The income effect inclusive of the behavioral response (11) is given by (12)

$$(12) \qquad dC_j/de_j = (\Omega - t_{ij}) w_j h_1 A_2 (1 + \omega_1 (dt_{ij}/de_j)) - (dt_{ij}/de_j) w_{ij},$$

which differs from the productivity effect by $-(dt_{ij}/de_j)[(\Omega - t_{ij}) w_j h_1 A_2 \omega_1 + w_{ij}]$. If the increase in local water quality reduces, water purification efforts, as is likely from (11), then consumption or income rises because more time is spent earning (the third term in (12)). On the other hand, the productivity-enhancing effect is attenuated, as embodied in the last term in (12).[5]

b. *Income and effort effects of As retention in a multiple-member household*

The above simple model illustrated some of the biological, informational and behavioral mechanisms by which arsenic retention variability is associated with productivity and observed economic outcomes. The model, however, is not well-suited for inferring the economic consequences of arsenic ingestion and retention in Bangladesh. First, in Bangladesh time devoted to activities associated with consumed water quality are typically done by non-earning women.[6] Second, earners are almost exclusively men, and if there are multiple men they are usually kin (brothers and/or sons of the household head). Third, 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.[7]

[5]Even if effort increases, the sign of the bias is ambiguous, as there are still two terms with opposite signs.

[6]In our 2008 data, 98% of household members who spend any time fetching water, gathering fuel (wood or dung), and/or cooking are women, almost 80% of whom are wives of the head. Of the women who participated in these activities, less than 3% also participated in the labor market inclusive of household earning activities such as farming, fishing or non-farm enterprises. Of all women aged 24-59, 94.9% report "housewife" as their primary occupation.

[7]59.7% of men in the labor force aged 24-59 are exclusively self-employed in our 2008 data.

In this section we examine the relationship between individual arsenic retention, productivity and consumption in a household context that incorporates these features of rural Bangladesh. We show that it is possible to identify both the existence of any home productivity effects of arsenic retention as well as the magnitude of market productivity effects with information on exogenous variation in individual-specific arsenic retention, on behavior associated with water purification, and on aggregate household expenditure, even if such expenditure does not value home-produced goods, as is typical in survey data.

We augment the previous model by adding multiple household members, distinguishing among earners and non-earners, and adding a home-produced good. In particular, 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. In addition to the arsenic retention and wage functions incorporated in the first model, we add a production function for the home good (13) and a home productivity 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) $H_{fj} = \mathrm{H}(A_{fj}, r_{ij})$ $\quad \mathrm{H}_1<0.$

To reflect that the earners are members of the same family lineage and the person producing the home good is a spouse (wife), and in accord with the literature on the inheritability of methylation efficiency, we assume that the methylation efficiency endowment has a common family component, so that $\mu_{ij} = \mu_j + \xi_{ij}$ for the earners, while the individual-specific μ_{ij} and μ_{fj} are uncorrelated.

The utility function for the multi-member household is

(15) $\mathrm{U} = (A_j, C_j, M_j) \mathrm{U}_1<0; \mathrm{U}_2,\mathrm{U}_3>0; \ \mathrm{U}_{11}> 0; \mathrm{U}_{22},\mathrm{U}_{33}<0,$

where A_j=average arsenic retention of family members. As before we assume for simplicity that arsenic is separable in (15), but we allow the home and purchased good to be non-separable ($\mathrm{U}_{CM} \neq 0$).

The budget constraint is given by

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

In this model, the spouse's retention of arsenic has no direct effect on income or the amount of the consumption good purchased. Arsenic may affect the wife's productivity in producing the home good, as reflected in (14). 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

under the same informational assumptions as in the first model as long as there is a time component to reducing arsenic ingestion through water purification effort. The FOC is

(17) $$U_A A_2 \omega_1 = U_M M_1$$ [8]

The model delivers the following proposition for identifying the existence of effects of the wife's arsenic retention on home productivity:

Proposition 2:

If H (skill) does not matter in producing the home good, the wife's individual arsenic retention will have no effect on water purification effort.

Proof: 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) $$dt_{fj}/d\mu_{fj} = \{A_3 H_1 [U_{MM} M_1 M_2 + U_M M_{12}]/\Phi^*,$$

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.

The model also delivers a prediction for how a change in arsenic retention experienced by household earners affects effort.

Proposition 3:

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).

Proof: The (common) effect of a reduction in arsenic among the earners on the wife's effort is:

(19) $$dt_{fj}/d\mu_j = N w_j \Omega h_1 A_3 U_{MC} M_1/\Phi^* = N w_j \Omega h_1 A_3 dt_{ij}/dF_j,$$

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, $U_{MC}>0$ (substitutes, $U_{MC}<0$). Note that the effect of arsenic retention among household earners on water purification effort depends on the number of earners in the household.

The estimates of the wife and earner A_{ij} effects on the wife's water purification effort, which identifies both the existence of home productivity effects and the relationship in the utility function between home and purchased goods, is informative about the household welfare losses from arsenic

[8]The FOC in the case where if the productivity effects of arsenic are known is $U_A A_2 \omega_1 = U_M M_1 - A_2 \omega_1 [U_M H_1 + U_C N \Omega h_1 w_j]$; again arsenic retention will be higher than optimal in the absence of this information.

contamination. With information on the exogenous variation in arsenic retention among earners it is possible to identify whether and how arsenic retention affects earnings productivity h_1. This is because the effect of variation in earners' A_{ij} on total household consumption C_j is $Nw_j\Omega h_1$, given (4) and (16).[10] The relationship between the effect of arsenic retention on consumption expenditures, however, may overstate or understate the effect on the welfare of the household, which also consumes household produced goods. If, for example, arsenic does not affect the home productivity of women, then the reduction in consumption expenditures overstates the welfare loss to the household, the extent depending on whether and by how much home and purchased goods are complements or substitutes. If increases in arsenic retention also affect home productivity, then the reductions in consumption expenditures more closely correspond to the welfare loss.

Finally, the model makes clear that the earnings effect on consumption from reduced arsenic retention is not the same as the effect of improving access to clean water - changing the quality of the household's local water source - on total household consumption expenditures. Estimates of the earnings and water purification effects of varying retained arsenic A_{ij} among both earners and non-earners, respectively, are also necessary for predicting the consequences of improving available water quality for the household on household income or consumption expenditures:

Proposition 4:

> *If reductions in arsenic retention A increase market productivity for earners and household productivity for non-earners, the relationship between water quality e_j and total household consumption expenditures will be less than the effect of reducing earner's arsenic retention A on earnings even if the effect of improving e_j on A is not known by households.*

Proof: The effect of a change in e_j on C_j is

$$dC_j/de_j = Nw_j\Omega h_1 A_2[1 + dt_{fj}/d\mu_j\omega_1/A_3] , \tag{20}$$

The first term in (20) is the direct effect of the common component of the earners A on expenditure. The second term, from (19) is negative if the decrease in arsenic retention from cleaner water augments productivity in home production.[11]

---

[10]We are assuming that As retention has no effect on the labor supply of (male) earners. We estimate this relationship below, and find that total annual work days (including days sel-employed) among males is unaffected by the amounts of their retained arsenic.

[11]If household's perceive that the water quality is improved, then the effect of improving water quality on household income will be smaller than given by (20), as then the second term becomes $\omega_1\{[(A_2)^2 U_{AA} + U_A A_{22}] + H_1[U_{MM}M_1M_2 + U_M M_{12}]\}/\Phi^*$, which is more negative than (20).

3. Method of analysis

Our main empirical objectives are to identify how concentrations of arsenic that result from the ingestion of arsenic-contaminated water causally affect health, human capital and productivity at the individual level, and arsenic-mitigating activities and income at the household level. Our first objective is to estimate the effect of As on measures of individual productivity and health. More formally, linearizing the productivity function (4), we seek to identify the parameter δ in:

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

where the l index identifies the individual's relationship with a family member, z_{ijlz} is a vector of observed exogenous attributes of the individual and household, u_{ijl} is an individual-specific error, u_j is a household fixed effect (reflecting, for example, the local health environment), and ε_{ijl} an iid error. As shown in equation (5) from the model, the parameter δ reflects not only the causal affect of As retention but also behavior; that is, 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 at the village level, such as the construction of a deep well to improve e_j, cannot be used as an instrument for individual arsenic stock because the new well will directly alter the allocation of time of all family members, for example, and thus affect labor market and other outcomes directly for any individual. Even randomized individual-specific interventions that reduce arsenic ingestion or increase arsenic retention cannot be used as instruments to identify δ. For example, the randomized distribution of nutrition supplements across individuals, which have been shown to increase arsenic metabolism and excretion (as in Gamble *et al.* 2007), would have direct effects on health and productivity in addition to any effects via the stock of arsenic. Such an intervention thus would not identify the arsenic-health or productivity mechanisms. 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 source of exogenous variation in the arsenic stock of the body, and hence the toxicity of any ingested quantity of arsenic, and form the basis for an IV estimation strategy. The evidence further suggests that, apart from the nutritional determinants of arsenic concentrations in humans, genetic

polymorphisms are a major cause of the substantial inter-individual variation in arsenic metabolism (methylation) 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 in urine. The enzymes that are required for this chemical process have also been identified and linked to four specific genes. 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. Engström *et al.* (2007) found that a significant part of the variation in urinary metabolites of arsenic in an Andean population is due to polymorphism in the AS3MT gene, as did Rodrigues *et al.* (2012) for a Bangladeshi population. Agusa et al. (2011) review 18 human case studies on this gene and find that two polymorphisms of the AS3MT gene were consistently related to arsenic methylation efficiency regardless of the populations examined for the analysis.

Variants of the glutathione-S-transferase (GST) enzyme have also been identified as catalyzing the methylation of arsenic. The glutathione S-transferase omega-1 variant of this enzyme is encoded by the GSTO1 gene and the glutathione S-transferase theta-1 variant is encoded by the GSTT1 gene. Both genes are also located on human chromosome 10. Kile *et al.* (1995) finds that individuals in Bangladesh possessing GSTT1-null genotypes (lacking the genes normal function) had significantly more arsenic in their toenails in contrast to GSTT1 wild-type (typical allele for a gene) individuals, after controlling for drinking water contamination. Also working in Bangladesh, McCarty *et al.* (2007) find that the GSTT1 wildtype modifies the risk of skin lesions among arsenic-exposed individuals. Chandra *et al.* (2011) find that persons in West Bengal (India) having null genotypes have significantly decreased urinary arsenic and increased clinical symptom scores relative to persons with GSTM1 or GSTT1 non-null genotype of the same arsenic exposure group.

Arsenic metabolism involves methylation to MMA and DMA by a folate-dependent process. Methylenetetrahydrofolate reductase (MTHFR) is a key enzyme in the metabolism of folate that has been strongly linked to both arsenic metabolism and toxicity in laboratory and animal studies. Ahsan *et al.* (2007) find that variations in the MTHFR and GSTO1 genes modulate health effects of arsenic exposure, noting that the consistent relationships of the risk of skin lesions with genotypes and diplotypes in the MTHFR and GSTO1 genes suggest the importance of genetic susceptibilities in their Bangladeshi study population. They estimate that the proportion of skin lesions in their study population 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, and there is some evidence of this in human populations. For example, in Chung *et al.* (2002) families from Chile were selected based on their long-term exposure to very high levels of arsenic in drinking water (735–762 μg/L). Each family consisted of a father, a mother, and two children. Urinary arsenic and its methylated metabolites for each participant was measured. The intra-class correlation coefficients showed that 13–52% of the variations in the methylation patterns were from being a member of a specific family. Obviously these correlations could arise due to similarity in diets and to common environmental factors. However, methylation patterns are correlated more strongly between siblings than between father–mother pairs within a household, after adjustment for total urinary arsenic, age, and sex. 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 and dietary choice. Using the notation of the model, the covariance between measured arsenic retention of two family members residing in areas m and n is

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

where μ_l is the common genetic component of μ_{ijl}. As seen in (22), as long as local prices 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. 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 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, we estimate the equation

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

where $\boldsymbol{N}_{ijk}$ is a vector of person-specific and endogenous family inputs, including the log of individual foods consumed, the log of smoking, and household's choice of a water source for cooking; and $\boldsymbol{\alpha}$ is a vector of coefficients. Equation (23) contains five sources of unobserved heterogeneity highlighted in the model: The first is the genetic component of arsenic methylation ability that is shared among a lineage or kin group μ_l, the second the individual-specific component of that ability. There is also a component that represents the unobserved exposure to environmental arsenic from local drinking and cooking water, e_k, a household error component u_j and an iid error term v_{ijk}.

To obtain consistent estimates of the $\boldsymbol{\alpha}$, 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 both water quality e_k and in individual methylation ability, 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, using as instruments for foods and water choice the village-level prices of foods P_{kx} interacted with exogenous person- and household-specific exogenous attributes, including own age and gender (which are also included in (23)) and 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.

The estimates of $\boldsymbol{\alpha}$ are of direct interest. First, we can quantify to what extent the individual consumption of fruits and green leafy vegetables, for example, which the literature suggests are a source of arsenic ingestion and also mediate the effects of arsenic ingestion via arsenic metabolism and excretion, affect arsenic concentrations in the body, given exposure to arsenic via water sources (impounded in the household fixed effect). For these estimates 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, so that it is not necessary to have the complete life

history of food intakes. The toenail-based measure mainly reflects arsenic ingestion and excretion in recent months, not years (Kile *et al.*, 2005).[12]

The residuals from (23) for genetically-linked but non co-resident family members can be used as instruments for A_{ijk} in (21) to identify δ. These 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. Our panel data contain many households residing in separate villages whose family links are identified. This is because of the panel design of the survey, which followed all household members who left the households in a prior round, and because 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. Because as part of the survey design we include both the originally-sampled households and any household containing any individual from the original households no matter where the location in Bangladesh, we have as much information on family members who remained together as we do for those who are no longer co-resident.

After estimation of (23), we compute $E(\mu_l | A_{ijl} - \boldsymbol{N}_{ijl}\boldsymbol{\alpha}$ for all members of lineage l) for all sampled respondents who have a non coresident family member residing in another village. The short-run nature of the arsenic concentration measure is again important - the fact that the separated family members once shared a common arsenic source should not be reflected in the toenail-based measure of arsenic as long as the family members divided at least a year prior to the survey. However, if there is measurement error in $\boldsymbol{N}_{ijk}$, then in lineages with small numbers of observations, this expectation will also be measured with error that may not be of the classical variety in subsequent estimation. If this measurement error is uncorrelated across member in a lineage, then it will be useful to estimate 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 $(dh_{ij}/d\mu_{ij})\text{var}(u_{ln}) + (dh_{ij}/dr_{ij})\text{cov}(r_{ij}, \mu_{ln})$, from which it can be seen that both the variance

[12]Kile *et al.* (2005) collected toenail and drinking water samples from forty-eight families (n=223) every three months over a two year period. The team created a weighted, lagged exposure variable including drinking water arsenic concentrations measured three, six and nine months before toenail collection. Their findings suggest that the drinking water concentrations at three, six and nine months contributed 69%, 14%, and 17%, respectively, to the arsenic detected in the toenail sample.

of the common genetic component of arsenic methylation var(u_{lh}) must be nontrivial and cov(r_{ij}, μ_{kn}) must be negligible for μ_{kn} to be a valid instrument for A_{ijk}. That is, identification of δ requires not only that the household components of the error terms across households not located in the same village are uncorrelated but that the genetic polymorphisms that regulate the efficiency of arsenic metabolism are unrelated to those that affect the outcomes of interest h_{ijk}.

Testing for this orthogonality among sources of genetic variation is critical for establishing the validity of the identification strategy based on within-family, cross village genetic correlations. We will carry out standard tests of overidentification. However, we will also use new data derived from recent investigations of the human genome that enable a direct test of this condition. 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 on the human genome suggests that they are unlikely to be correlated. However, we will carry out and report the results from Bonferroni-adjusted tests (Sidak, 1967) of multiple pairs of correlations between the arsenic genetic polymorphisms and the human capital polymorphisms using data from a sample of genetically-linked individuals whose genome has been well characterized as part of the International HapMap Project.[13]

4. Data and Measurement of Arsenic Concentrations

a. Survey design and sample size

The data set we use consists of information on 13,258 individuals residing in 2,480 households from the 2007-8 round of a panel survey that started in 1981-2 in 15 villages in rural Bangladesh. The first survey was obtained using a sample that was meant to be representative of the rural population of Bangladesh (Ahmad and Hassan, 1982). The second round of the survey in 2002-3 included all of the individuals in the 1981-2 households residing in 14 of the 15 original villages, wherever they were located in Bangladesh, plus all of the individuals residing in any new households formed since the original round. Less than 3% of non-deceased individuals from the first round were not found and included in the survey second round. In addition, at the time of the second round of the survey a new random sample of households in the original 14 villages were surveyed, and in the latest 2007-8 round all of the individuals

[13]The International HapMap Project has developed a haplotype map of the human genome, the HapMap, which describes the common patterns of human DNA sequence variation. The HapMap is a key resource that is used by researchers to identify genes affecting health, disease, and responses to drugs and environmental factors. We describe the data we use from this project below.

from the 2002-3 round and any new household members were included as respondents. Attrition for this round was less than 8%.

Because of the panel survey design, which tracked all individuals who were ever sampled regardless of location, kinship relationships with other sampled individuals but residing in different households and villages are identified. The spatial separation of kin who once co-resided results from (marital) migration. In the 2007-8 round of the data the number of villages represented in the sample had grown because of migration from 14 in 1981-82 to 612. The information on family links among spatially-separated respondents, as noted, is key to the identification of the causal effects of arsenic ingestion and retention, given the genetic component of arsenic methylation.

The survey data provide more detailed information at the individual level than do most large-scale surveys. The data include individual-specific information on food intakes, observed by investigators over a 24-hour period. This information is required to carry out our estimation strategy in which we eliminate diet as a source of retained arsenic variation. There is also a wide variety of individual-level productivity and health measures, including information on anthropometrics, morbidity, occupations, household activities and labor force participation. In the last round of the data measures of physical capabilities, measured by pinch strength, and cognition, measured by an abridged version of the Raven's Colored Progressive Matrices exam, were also obtained for every respondent meeting minimum age requirements.

b. Measurement of respondent arsenic concentrations

A key component of this study is the measurement of the arsenic concentrations of the respondents. We collected from each respondent aged eight years and above clippings from all ten toenails, providing a basis for a contemporaneous measure of arsenic ingestion for 7,356 individuals. To extract information on arsenic concentration from toenails, trace metal analysis using inductively-coupled plasma mass spectrometry (ICP-MS) must be used after preparation, including cleaning with solvents and digestion in acid.[14] Toenail clippings are the preferred biomarkers for measurements of arsenic ingestion in every case in which exposure lasts more than a few months, and is particularly preferred when a significant share of arsenic exposure occurs through food consumption (Kile *et al.* 2007). Kile *et al.* (2005) also demonstrate that arsenic concentrations in toenails vary with genetic polymorphisms known to affect methylation efficiency, using a sample of households from Bangladesh.

[14]Appendix A provides a detailed description of the methods that were used at the Environmental Chemistry Facility at Brown University in preparing the toenail samples and at the ICP-MS Laboratory at the Graduate School of Oceanography at the University of Rhode Island where the trace metal analysis was performed.

Because of the considerable expense of the laboratory analyses and preparation, we assayed only a subset of the toenails. In particular, we confined the analyses to pairs of households containing related kin residing in different villages that have at least two, same-sex respondents within each pair. Although it is useful to have this information for all respondents for which we have these biomarkers, our identification strategy, as noted, requires data on arsenic concentrations for closely-related relatives living apart. Because of budget limitations, the number of over-eight respondents for whom we have arsenic concentration (*As*) measures is 4,260. Due to the sampling strategy used for the laboratory analyses, the sample of respondents with the *As* information slightly over-represents married women, due to their relatively greater mobility, and is somewhat older than the sample of respondents with unanalyzed toenails. Neither average schooling levels nor landholdings, however, differ between the two subsamples. In all of our empirical analyses using this information, we will separately analyze males and females and control for age (and landholdings).

As part of the assessment of the laboratory methods, a sample of 25 respondents (graduate students residing in Rhode Island) also provided toenail clippings. These samples provide not only a basis for assessing the capabilities of the laboratory methods used to measure low levels of *As* concentrations, but also a benchmark with which to compare the levels measured from the Bangladesh sample. The analyses indicate that the concentrations of arsenic in the Bangladesh respondents are quite high, vary considerably, but are spread across all economic groups. In the US graduate student sample, average *As* concentrations are 78.3 parts per billion (ppb), with a standard deviation of 46.6 . In the sample of Bangladesh respondents the average concentration is 1,353 ppb, with a standard deviation of 1,894. Figure 1 provides the frequency distribution of the arsenic concentrations measured in the two samples, which show the substantial contamination of the Bangladesh respondents - 90% of the Bangladesh sample have *As* concentrations greater than the highest value found in the US sample, and over a third have concentrations exceeding 1,000 ppb. Figure 2 shows that arsenic contamination is not confined to the rural poor in Bangladesh - levels of *As* concentrations are actually slightly more elevated among households with larger owned landholdings.

c. Analysis sample

For the analysis of the effects of individual *As* concentrations on nutritional status, capabilities and earnings, we use a sample of adult respondents aged 18-59 for whom we have toenail clippings and have immediate family members, also with *As* samples, who reside outside of their village. For this subsample, we have the *As* concentration measure for 52.3% of female and 44.5% of male respondents, the slight imbalance due to the oversampling of toenail samples for split family members and the higher mobility of

women. Respondents in the subsample reside in 465 villages. Based on the kinship relationships, we constructed 583 lineage groups - - respondents living in different villages who are either a sibling or a parent-child pair.

Table 1 provides average individual-specific *As* concentrations, food intakes and outcome measures by gender for the subsample. As can be seen, the level of retained arsenic is approximately the same as the age-unrestricted sample, with women having concentration levels that are slightly higher than that of men, by 6.5%. This difference is not statistically significant at the 10% level. Women, however, also appear to consume less foods overall, and thus potentially less contaminants. And men smoke on average almost seven times more cigarettes per day than do women. In the table we also see that women do less well on the Raven's test, are less strong, have less schooling, spend less time in the labor market and are less likely to operate a business, but these statistically significant differences are not necessarily attributable to the small differences in retained *As*.[15]

The data also indicate that retained individual *As* is significantly correlated even among separated family members. In a fixed-effects regression estimated on the subsample of respondents who had left the original fourteen villages the set of 14 dummy variable coefficients associated with the original villages was statistically significant at the .01 level, consistent with equation (22). As one diagnostic for whether our instrument removes the village-level source of spatial correlation among separated kin, we will estimate the same regression using the estimated *As* residuals from (23). Our procedure should eliminate village-level sources of spatial covariance as well as any kin-based persistence in food habits or preferences for water quality associated with choice of water source.

5. Diet, Water Source and Arsenic Concentrations

We first estimate equation (23), the relationship between food intakes, water source and arsenic concentration, using the individual-level information on food consumption, divided into seven food groups (grains, pulses, green vegetables, other vegetables, tubers, fruits, and meat, fish and dairy); information on smoking (number of cigarettes per day); and information on the water source used for cooking, coded as a binary variable if a non-tubewell source was used. Over 26% of households used water from non-tubewell sources for cooking. In contrast, more than 97% of households obtained drinking water from a tubewell so that there is too little cross-household variation in this variable to obtain an estimate of the effects on arsenic retention of switching sources of drinking water.

[15]For example, women have less physical strength in all human populations (Pitt *et al.*, forthcoming).

Avoiding tubewells as a source of cooking water appears to be associated with effort.[16] The distance to the water source for cooking is a statistically significant (.03 level) 15% higher for users of non-tubewell sources and time spent fetching water in such households is a statistically significant 19.6% higher (.037 level). Consistent with the fact that over 30% of households boil water that is not obtained from a tubewell, in households using non-tubewell water for cooking, time spent fetching fuel is also a statistically significant 32.1% higher (.001 level) than in households who use tubewell water for cooking.

As noted, the toenail-based *As* concentration measure represents arsenic retained in the body from arsenic ingestion over the past three months, while the food intakes are measured in a 24-hour period. Both the outcome and input variables thus are short-term, but the food intake variables measure with error the food consumed over the period relevant to the concentration measure. Our instrumental-variables method should eliminate this source of bias, along with the biases due to the existence of unobservables that affect the choice of foods.

Table 2 reports OLS and IV estimates of the diet-arsenic production function. All food variables and the quantity of cigarettes smoked are expressed in logs, as is the concentration of arsenic. While the signs of the OLS and IV coefficients are identical for all but grains (which is, however, the largest single food item), the OLS coefficients for all endogenous variables are biased towards zero. The Anderson test for underidentification indicates strong rejection of the underidentification null. The hypothesis that the set of OLS and IV coefficients are identical is also rejected (p=.0015). The estimates of the gender effect, using either estimation method, indicate that women at age 30, net of dietary intakes, retain 5.5% less arsenic in their bodies than do men.[17] That in our sample, on average women have more *As* concentrations than men thus appears to be because in part women consume different diets.

The IV estimates indicate that the staple of rural Bangladesh diets, grains (principally rice, a food that uses large amounts of water for cooking), 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.[18] The point estimate indicates that a one-standard deviation increase in grain

[16]We cannot know whether such effort reflects attempts to reduce arsenic ingestion, though we find below that use of the alternative cooking water source does reduce retained *As*.

[17]This result is consistent with the medical literature indicating that women methylate more more efficiently than do men as a consequence, in part, of the protective effect of estrogen (Lindberg *et al.* 2007).

[18]Irrigation with arsenic-contaminated water particularly affects rice. This is partly because of the large amounts of water used to irrigate rice and partly because the form of arsenic present in a flooded field is

consumption increases arsenic retention by 12.6%. Smoking also statistically significantly increases arsenic retention, a finding consistent with medical studies.[19] The point estimates indicate that the cessation of smoking would lower retained arsenic by 4%. The consumption of three food groups, however, significantly decreases retained arsenic in the contaminated water environment of Bangladesh - tubers, meat and green vegetables. Recall that green vegetables, based on evidence obtained from the randomized distribution of folate supplements, given arsenic ingestion, lower arsenic retained in the body because of increased methylation.[20] And there is also recent evidence that tubers also reduce *As* based on population samples in Bangladesh (Pierce *et al.*, 2010), with consumption of roots and gourds being negatively associated with the appearance of skin lesions. Finally, the estimates indicate that there is a substantial payoff from shifting the source of cooking water from wells in terms of arsenic retention - switching from wells to obtain water for cooking evidently decreases retained *As* by 18.2%, a result which is statistically significant at the .03 level, one-tailed test.

6. The first-stage equation

As noted, we use the estimated residuals obtained by subtracting from the measured *As* of all respondents that part predicted by own consumption, household water choice and the village fixed effect using the estimates reported in Table 2 to form respondent-specific family lineage measures of arsenic retention that contain genetic but not behavioral components. These lineage "endowments" are then used as instruments to predict respondent arsenic retention based on the genetic linkages in arsenic methylation.

---

the form that is most readily available to plant roots (Brammer and Ravenscroft, 2009). Rice is also much more efficient at accumulating arsenic into the grains than other staple cereal crops, irrigated or not (Bhattacharya *et al.*, 2012). In addition, rice readily absorbs arsenic when boiled in contaminated tubewell water. Huq *et al.* (2009) report that even if an uncooked rice sample did not contain any detectable amount of arsenic, the cooked rice (*bhat*) contained a substantial amount of the element arsenic when it was cooked with arsenic-contaminated water and Mahal *et al.* (2010) find that rice cooked in surface water contained less arsenic in Bangladesh.

[19]Chen *et al.* (2007) and Lindberg *et al.* (2010) have demonstrated that smoking is associated with poorer methylation capacity. The effect of smoking may be related to competition between arsenic and some of the many chemicals found in cigarette smoke for common detoxification pathways (Hopenhayn-Rich *et al.*, 2006). Studies of Bangladesh sub-populations for whom measures of the products of the methylation process, arsenic metabolites, were obtained using spectrometry on biomarkers, have found that smoking interfered with arsenic methylation capacity but was not a direct source of arsenic (Kile *et al.*, 2009).

[20]On the other hand, green leafy vegetables in Bangladesh are strong accumulators of arsenic, much more so than fruity vegetables like tomato, gourd, or eggplant (Farid *et al.*, 2003). Arum leaf, a popular and widespread green vegetable, has the highest arsenic load of any foodstuff in Bangladesh tested by Huq *et al.* (2006).

Note that because we have eliminated any fixed effects associated with villages, we require that at least two different lineages reside in a village.

We first assess if our method of eliminating the environmental and behavioral components of measured *As* successfully eliminated the spatial correlation in retained arsenic levels. We regress the measured *As* of those respondents who had left the original 14 villages on their age, age squared, gender and the value of their households's landholdings and a set of 14 village dummy variables corresponding to their village of origin. The origin-village fixed effects are highly jointly significant ($F(13, 505)=46.33$) and explain 56% of the total variance in arsenic retention across the sample of leavers. Spatial and behavioral correlations in *As* are evidently high in our sample, perhaps due to selective migration. We then replace the dependent variable with the residual measure of *As*. Using the same sample and specification, the set of origin-village dummy variable coefficients is no longer jointly statistically significant ($F(13, 505)=0.71$), and explains just 2.2% of the total variance in the residual measure.

Having successfully eliminated the environmental sources of the correlation in *As* among family members living apart using the residual method, the next question is whether and how the family-based residual measures explain the variation in actual arsenic retention across individuals. Genetic theory suggests that the functional form of the expected relationship between an individual's genetic ability to methylate arsenic and that of their family members is non-linear. In particular, the functional form of the first-stage equation is likely to be best described by a polynomial relationship if the inheritability of arsenic methylation efficiency is both polygenic and epistatic, as the genetic model with these attributes generates a nonlinear relationship between the phenotypes of any one brother and the phenotypes of his siblings that can be approximated by polynomials. One example of polygenic and epistatic inheritability is when there are three genes that determine a particular characteristic and the alleles of one gene must be of a certain type for there to be effects of the other two genes.

The literature leaves no doubt concerning arsenic methylation's polygenic nature. Three genes have been identified that are closely related to arsenic methylation efficiency (AS3MT, MTHFR, and GSTO1), discussed in more detail in Section 2 and Appendix Table A. Epistatis, or gene-gene interaction, in complex metabolic mechanisms, such as the methylation of arsenic, is considered likely as they require many enzymes that typically function together, and the interactions inherent in these biochemical relationships play a key role in determining epistasis. Lehner (2011) suggests that the simplest molecular mechanism that can cause epistasis between two genes is if their two protein products directly interact (p. 324). The AS3MT, MTHFR, and GSTO1 genes strongly identified as sources of variation in arsenic methylation

efficiency in human each regulate an enzyme required in the process and interact with other enzymes (Vahter 2000). [21]

Figure 3 plots the locally-weighted estimated coefficients from a regression of the endowment residual of a sample respondent on his or her average lineage residuals, by the level of those residuals, from our sample of respondents aged 18-59. As can be seen, the relationship suggests a quadratic form. To approximate this relationship we thus use both the level and the square of the average lineage residuals as instruments for a respondents log *As* retention. Table 3 reports the linear and quadratic specifications for the first-stage equation that we use in all of our subsequent IV estimates. The addition of the squared term adds explanatory power, and the two lineage variables are jointly significant at the .0025 level. The set of additional variables are included because they will be used in all of the second stage equations. Of these, only the value of the household's landholdings is statistically significant, indicating that wealthier rural households, net of their genetic tendencies to methylate, have slightly higher levels of retained arsenic.

The first-stage equations are estimated including both men and women. The hypothesis that all of the coefficients are the same for males and females cannot be rejected for either specification. As a consequence, and because we also could find no differences in the first-stage estimates across age groups, we will use the same first-stage equation for all of our estimates, including those in which we stratify by gender and/or age, using limited information maximum likelihood (LIML).

7. Retained Arsenic and Individual Performance

a. Arsenic and Cognitive Performance.

The first column of Table 4 reports OLS estimates of the relationship between performance on the Raven's Colored Progressive Matrices Test and the log of respondent's retained *As* for men and women aged 17-59. The estimate suggests that retained *As* and test performance is statistically significantly negatively correlated. But is the relationship causal? The fact that the household's landholdings is positively associated with the test score suggests that there may be a nutritional or behavioral component to the relationship. For example, it is well known that schooling attainment has some effect of Raven's performance. In the second column we report the estimates using two-stage least squares, using the first-stage equation reported in Table 3. The estimate of the *As* effect is now larger and still statistically

[21] Argos (2011) has found strong empirical evidence of gene-gene interactions for a set of 10 SNPs associated with arsenic methylation using data on 1,689 individuals from rural Bangladesh.

significant. The point estimate is large suggesting that a one standard deviation decrease in arsenic retention would increase performance on the test by one full correct answer, an increase of 24%.

The Wu-Hausman test indicates rejection of the hypothesis that *As* is exogenous, and the standard diagnostics indicate rejection of the hypothesis of weak instruments. In particular, the value of the Cragg-Donald *F*-statistic of 24.8 is well above the critical Stock-Yogo (2001) values for determining bias in the instrumented variables.[22] The Hansen *J* overidentification test also indicates non-rejection of the hypothesis that one or the other of the instruments are excludable. To assess whether this test has power, we re-estimated the test score equation including the actual lineage average *As* in addition to the residual measures in the first stage specification. The resulting estimates of the second stage are reported in the third column of Table 4. The Hansen *C* test indicates, as expected, rejection of the hypothesis that the family *As* "instrument" is excludable, while still indicating non-rejection for the residuals-based instruments.

b. Arsenic and Physical Strength.

Table 5 reports OLS and LIML estimates of the relationship between retained arsenic and a measure of physical strength - performance on a standard pinch test. Each respondent was asked to pinch a dynometer with each hand three times. We use the sum of the pressure exerted in all six tries (in kilograms of pressure). As for the cognition tests, the OLS estimate of the effect of retained *As* is substantially underestimated. The OLS estimate is not statistically different from zero, while the LIML estimate which accounts for endogeneity is statistically significant at the .05 level. The point estimate indicates that a one-standard deviation increase in retained arsenic reduces performance by over 6%. Also, as for cognition, while the point estimate of the effect is smaller in absolute value for women than for men, the differences by gender are not statistically significant.

c. Are the performance results spurious?

1. Arsenic and schooling attainment by cohort. One reason that we may find that use of our instruments, which rely on the genetic correlation in methylation within family lineages, results in significant effects of arsenic retention on cognitive and physical performance is that methylation genes are simply negatively correlated with genes determining inherent cognitive ability and strength, which are also inheritable. To assess whether our results are spurious we first examine schooling attainment by cohort. We should expect that those individuals with lower cognitive performance, whatever its origin, will choose less

[22]For example, the critical 5% *F*-value for two instruments when weak instruments are defined so that a 5% hypothesis test rejects no more than 15% of the time is 11.59. (Source: Table 1 from Stock *et al.* (2002).

schooling and thus those respondents with higher predicted *As* concentrations should have lower schooling attainment.[23] However, because the shift to arsenic-laden well water started in the late 1970's, older respondents should not have been affected by arsenic poisoning when they made their schooling decisions. Thus, we have a cohort test of spuriousness - if we find older cohort's completed schooling is as related to current arsenic retention as that of younger cohorts, this would suggest that *As* is simply correlated with fixed attributes that affect cognition, for example due to a correlation in ability to learn and methylation ability.

To assess first if the current cognitive performance of the old and young are equally affected by their current arsenic retention, we re-estimated the Raven's test equation, using LIML, for the age group 18-34, who were likely consuming arsenic-tainted water throughout their entire lives, and the cohort aged 35-59, most of whom were not likely consuming arsenic-contaminated water from wells as children. Table 6 reports the estimates for both age groups. As can be seen, the estimates of the *As* effects on cognitive performance are no lower for the older cohorts. We cannot reject the hypothesis that the estimates are identical by age group - current arsenic retention evidently affects current cognitive performance for both old and young - and equally by gender within age groups.

When we examine schooling, however, the results are strikingly different. The relationships between current *As* retention and completed schooling by cohort and gender, unlike for the test score results, appear to reflect the historical conversion to tubewells as sources of water. Table 7 reports LIML estimates of the relationship between arsenic retention and number of years of schooling for two age groups - those aged 23-34, who have all completed their schooling and were likely consuming large amounts of arsenic during their school decision years, and those aged 45-64, who were not likely affected by arsenic while of school age - and by gender.

As can be seen, for both men and women the negative relationship between current *As* retention and completed schooling is substantially stronger for the younger cohort. However, for women, the estimates are not statistically significant for either cohort, and are much smaller than those of men. Among the men, current arsenic retention and schooling attainment are strongly and statistically significantly negatively correlated for the young cohort, while as expected for the older cohort of men, whose contemporaneous relationship between *As* retention and cognition is identical to that of the young, current

[23]Choice of schooling should depend on comparative advantage in skill versus brawn. Our estimates indicate that retained *As* lowers cognitive performance substantially more than it decreases physical strength.

arsenic retention and schooling attainment are unrelated. This difference by cohort is statistically significant. To see if these results for men are sensitive to the choice of age group, and given that the historical switch to tubewell-sourced water was likely gradual, we also used local-IV to estimate the relationship between current arsenic retention and schooling attainment by age across the cohorts aged 23-59. The lowess-smoothed local estimates by age for the men, along with the relevant 95% confidence bands, are shown in Figure 4, which, as expected, display a monotonic decrease in the negative relationship as the year of birth retreats in time.

2. Results from genomic data. The schooling attainment results by cohort suggest that the relationship between contemporaneous arsenic retention and cognitive performance does not simply reflect a correlation between innate cognitive skill and the ability to excrete arsenic. In this section we use genomic data to test directly whether there is a genetic link between genes reliably associated in the genetics literature with both cognitive ability and strength and those demonstrated to be associated with arsenic methylation.

The basic genetic variations examined are single nucleotide polymorphisms (SNPs), which are DNA sequence variations that occur when a single nucleotide (A (adenine), C (cytosine), T (thymine), or G (guanine)) in the genome sequence is altered. For example, a SNP might change the DNA sequence AAGCCTA to AAGCTTA. In this case there are two alleles : C and T. Almost all common SNPs have only two alleles. For a variation to be considered a SNP, it must occur in at least 1% of the population. SNPs, which make up about 90% of all human genetic variation, occur every 100 to 300 bases along the 3-billion-base human genome.[24] For a pair of SNPs, the combinations of alleles define haplotypes.

Two SNPs are said to be in linkage disequilibrium (LD) when alleles at two or more loci (places on the DNA sequence, such as a SNP) 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, on the order of tens or hundreds of kbs (kilo base-pairs). LD between two SNPs can be quantified as the correlation between alleles across population chromosomes.

One standard measure of this correlation is the square of the correlation coefficient, which may be expressed as functions of the allele frequencies of the two SNPs. Consider the haplotypes for two SNPs 1 and 2 with two alleles each. Denote a variable X_A to be 1 if the allele for the first SNP is A and 0 if the

[24]Two of every three SNPs involve the replacement of cytosine (C) with thymine (T). Over 50 million SNPs are listed in the dbSNP respository (http://www.ncbi.nlm.nih.gov/projects/SNP/).

allele is a, and a variable X_B to be 1 if the allele at the second locus is B and 0 if the allele is b. A haplotype is the pair defined by the realizations (X_A,X_B). The linkage disequilibrium coefficient D_{AB} is

(24) $$D_{AB} = P(AB) - P(A)P(B),$$

where P(A) and P(B) are the frequencies of A and B, respectively, and P(AB) is the frequency of haplotype AB. The linkage disequilibrium correlation coefficient between X_A and X_B variables is:

(25) $$r_{AB} = D_{AB}/P_A(1 - P_A)P_B(1 - P_B).$$

Shifman *et al.* (2003) have examined the relationship between r-squared and base-pair distance across the human genome. Mean r-squareds are close to 0.1 for SNPs even 160 kilo-base-pairs (kbp) (out of 300 billion base-pairs) away and fall to near zero at a distance of 500 kbp.[25]

As discussed, a number of SNPs have been related to variations in the ability to methylate arsenic both from knowledge of the biochemical processes involved in arsenic methylation, and from associations in the data. Table A in the Appendix describes five SNPs related to arsenic methylation, and the name and metabolic function of the genes on which they are located. A review of the literature suggested a set of SNPs for intelligence (five SNPs), body mass and height (eight SNPs), and muscle development and strength (five SNPs) used in the LD analysis. Appendix B briefly describes these 18 SNPs and the criteria we used to select them based on a review of the literature.

The data set we use is HapMap Release 22, which uses single nucleotide polymorphisms (SNP) information from HapMap phase II that consists of 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[26] 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 (González-Neira *et al.*, 2006; Xing *et al.* 2008).

---

[25]Linkage disequilibrium is gradually broken down by recombination and gene-conversion. Recombination is the crossing-over of chromosomes during the process of sexual reproduction in which DNA is exchanged between a pair of chromosomes. Two genes that were previously unlinked, being on separate chromosomes, can become linked because of recombination. Gene-conversion is the non-reciprocal transfer of genetic information that alters the sequence of the recipient DNA. It is one of the ways a gene may be mutated. In rare cases, LD may extend for longer distances, for example, when the effective population size is small or in populations that have undergone recent admixture, neither of which characterize the Bangladeshi population.

[26]There are three other HapMap populations: Yoruba from Ibadan, Nigeria; Japanese from Tokyo, Japan; and Han Chinese from Beijing, China.

The r-squareds were estimated by maximum likelihood using the EM algorithm. Table 8 presents χ-statistics for pairwise tests of linkage disequilibrium obtained by computing the r-squared for each of five arsenic SNPs with each of the set of SNPs associated with intelligence, body mass and height, and muscle development and strength. Thus, these are tests of whether genetic variation in the ability to methylate arsenic, as determined by a set of SNPs, is correlated with genetic variation affecting important aspects of cognitive and physical functioning, as determined by a different set of SNPs.[27] The last column of Table 8 presents critical-rejection *p*-values using the Holm-Sidak Bonferroni-type adjustment for multiple comparisons. The *p*-values are 0.54, 0.23, and 0.80 for intelligence, body mass and height, and muscle development and physical strength respectively. For the full set of SNPs, the p-value is 0.49. None of the corrected *p*-values are close to common levels of statistical significance.[28]

7. Retained Arsenic and Health

In this section we use the same estimation methods to assess if an individual's retained arsenic affects health using conventional morbidity and anthropometric measures. Specifically we look at the relationships between retained *As* and the probability of having any one of a standard set of morbidity symptoms in the week prior to the survey for the same sample. We also examine how body mass (BMI) and retained arsenic covary. The morbidity measure we use is based on a set of illness symptoms that respondents self report, including headaches, diarrheal symptoms, fever, and coughing. Over 25% of the sample respondents aged 18-59 reported having had at last one of these symptoms in the reference period.

The first two columns of Table 9 report the probit and LIML probit estimates of the relationship between retained arsenic and the morbidity measure. Neither the probit nor the LIML probit estimate, which accounts for the endogeneity of retained arsenic, indicates that arsenic in the body manifests itself in standard morbidity symptoms. For BMI, while the OLS estimate of the association between BMI and *As* is negative and statistically significant, when the endogeneity of *As* is accounted for, the coefficient is no

[27]The chromosome and base-pair locations of all the SNPs are presented in Appendix Table B. The only SNPs that share a chromosome with any of the arsenic SNPs are rs821616 (intelligence) and rs211683 (body mass). Both are located on chromosome 1 as are two of the arsenic SNPsS (rs1801133 and rs1801131 on the MTHFR gene), but both are more than 500 kbp away. The r-squareds that we estimate for these sets of SNPs are consistent with the location data.

[28]To check if the tests of linkage disequilibrium have power, we estimated r-squareds and 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. The p-values were less than 0.001 in 3 out of 5 cases.

longer significant. The point estimates for BMI yielded by either estimation method are in any case very small, suggesting that a doubling of retained arsenic would decrease body mass by less than 1.7%. The reduced capabilities caused by arsenic do appear to be "hidden" - respondents with high levels of retained arsenic are neither experiencing any more illness symptoms nor are they smaller than respondents with low levels of retained arsenic.[29]

8. The Behavioral and Economic Consequences of Retained Arsenic

a. Individual effects: occupational choice, entrepreneurship and time worked.

In this section we estimate the effects of retained arsenic on occupational choice, entrepreneurship and labor supply using the subsample of young men aged 24-34. We focus on this age group because we expect that the selection of career paths occurs early in the life-cycle and, as indicated by the schooling attainment findings, older respondents were likely unaffected by arsenic in their water sources while young. To characterize occupation we used the information on the primary occupation reported by respondents and classified the respondents as being in a skill occupation according to whether it seemed reasonable skill and/or decision-making were important in the occupational category.[30] Of the 65 categories for workers, we categorized skill workers as those in the professions (teacher, doctor, government administrator) or in business management including farmers (but excluding a farm worker) and shopkeepers. For entrepreneurship, we used survey information on the nonfarm enterprises of individuals and categorized any respondent who reported having received nonfarm business income in the year prior to the survey as an entrepreneur. With these definitions, in this age group 49.2% of the men were in skill occupations and 19.2% were running, or were at least a residual claimant of, a nonfarm business.[31]

The first two columns of Table 10 report the probit and LIML probit estimates of the determinants of being in a skill occupation. The OLS estimate suggests that arsenic retention and being in a skill occupation are unrelated. However, the estimation procedure taking into account the endogeneity of retained arsenic indicates that increases in the amounts of arsenic stored in the body reduce significantly the

[29]Skin lesions (keratosis) and discoloration (melanosis) are visible physical symptoms of retained arsenic, but these in many cases take years to show up, and are only imperfectly correlated with the amounts of retained *As*. They are not included among the symptoms listed in the survey instrument.

[30]The occupational codes used in the 2008 round are identical to those used in the 1996 Matlab Health and Socioeconomic Survey.

[31]We do not examine the determinants of occupation choice or entrepreneurship for women, as only 6.6% of women spent any time in the labor force in the year before the survey and only 1.2% of women reported having a nonfarm business.

probability of being in such occupations, a result consistent with the impairment of cognitive functioning and subsequent reduction of schooling. The point estimate is relatively large, indicating that cutting by half the average levels of arsenic in this sub- population would increase the proportion in skilled occupations by 12.2 percentage points, or by 24%.

We find a similar result for entrepreneurship in columns 3 and 4 of the table - retained arsenic significantly reduces entry into entrepreneurship, with the effect substantially underestimated when the endogeneity of retained arsenic is not dealt with. The LIML probit point estimate suggests that 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%.

Finally, while elevated levels of retained arsenic evidently are causally associated with a male labor force of reduced skill and with fewer entrepreneurs, the quantity of work carried out by men with higher levels of retained arsenic is no less than that of workers with lower concentration levels of arsenic. Losses in income from arsenic poisoning are thus going to come from lost productivity rather than from reduced work time.

b. Individual earnings and household income losses.

As indicated in the multi-member household model, it is possible to estimate the effects of arsenic retention on market earnings using information on the individual-specific measures of arsenic retention among earners and total household expenditures. The major advantage of using the household consumption expenditure approach is that we can obtain estimates of earnings losses using all sample households that contain any men of labor-force age regardless of whether they earn wages or are self-employed. We thus avoid selectivity bias that would afflict estimates of earnings effects using only the subsample of wage earners, who represent less than half of prime-age males in the labor force in Bangladesh.[32]

The household expenditure equation that we estimate is given by

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

where $A_{ijm(f)}$ =log arsenic retention of prime-age men (women), $N_{m(f)}$=number of prime-age men (women), and $\mathbf{Z}$ is a vector of control variable such as the 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

[32]Of course, for identification we use lineage-based instruments that necessitate that some family members leave their origin villages. Given that almost all daughters leave their origin village to marry, in the context of Bangladesh this is not nearly as strong a restriction as the requirement that a man work for wages.

do not participate in the labor force. The model indicates that $\beta_2 = w_f \Omega h_1 < 0$, which is the earnings loss for a male earner from an increase in retained arsenic.

Equation (26) cannot be estimated using OLS, as we have seen that levels of arsenic depend on the composition of expenditures so that the A_{ij} will be correlated with the error term. If every household member belonged to a distinct lineage group and we had information on the arsenic levels of their non-coresident kin, estimation of (26) for a multi-member household using IV would be straightforward extension of the approach we have taken for an individual's measured productivity. However, this is not the case. First, a subset of the members of the household are members of the same lineage. As depicted in the model, our lineage-based instrument can be used to predict retained arsenic for all household members within the same lineage - the instrument is the common component of their ability to methylate arsenic. We thus have to assume that the effect of variation in retained arsenic within gender and age groups is the same for all members of the lineage in the household. This does not seem to be a strong restriction, as we can divide up household members into categories by age and gender. What is necessary for identification is that 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. For example, if all households had a prime-age male head, his brother and a sister ($N_m=2$, $N_f=1$), we could not separately identify all of the β's because we would only have a common prediction for retained arsenic for all of the members in each household. This exclusivity condition, however, is easily met in the data: few prime-age sisters or daughters of heads co-reside. This enables identification of β_1 and β_2. Conversely, when we have lineage instruments for wives, because wives do not co-reside with their father or brothers, and as long as there are some households where wives do not have prime age adult sons, β_3 and β_4 are identified. We do not, of course, restrict the sample to these subsets of households. The restriction is that the effects of retained arsenic on aggregate household expenditures is the same by gender and age-group regardless of household composition.

A second limitation of our lineage-based instruments is that we only have lineage instruments for members of households who were in the prior rounds of the survey. We thus do not have lineage instruments for all members of households. For example, in a household containing two married brothers who are of earning age ($N_m=2$, $N_f=2$), we might have a lineage instrument for the two brothers, because they had sisters in an earlier round who left their household and village to marry, but not their wives, who come from households outside of the original set of surveyed households. For such households we can estimate β_1 and β_2, under the assumption that the effect of retained arsenic on earnings is the same for each brother. We cannot, however, estimate β_3 and β_4, so we need to exclude the wives' arsenic levels from the

equation. The absence of lineage instruments for all household members thus requires an additional orthogonality restriction - that there is not assortative mating with respect to methylation genes. The exclusion of spouse arsenic levels will then not impart bias in the estimates of β_1 and β_2 (or β_3 and β_4 if the spouse is male) - the instruments for the brothers' retained arsenic, based on the retained arsenic endowments of their non coresident kin, will be uncorrelated with the wive's retained arsenic.

How plausible is it that there is not marital sorting by the ability to methylate? We expect that marital sorting will be attentive to the arsenic contamination of water sources, to the extent these are identified, so area-specific arsenic contamination is likely to have conditioned marriage sorting in Bangladesh in recent years. But in the absence of strong signals of individual arsenic retention for persons in a given environment, it does not seem likely there will be sorting on the relevant genes. And our estimates suggest that obvious outward signs of retained arsenic - stature and conventionally-measured morbidity - are not useful for identifying individual-specific arenic retention.

We can carry out a test of marital sorting by arsenic methylation ability, making use of the survey information on age at marriage. 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. The first column of Table 11 reports fixed-effects village estimates from a regression of the husband's log level of retained arsenic on that of his wife from a combined sample of couples who either married before 1981 or after 1990. Not surprisingly, given the couples likely consume and cook from a common water source and consume similar food, the coefficient is large and statistically significant. However, when we add an interaction term allowing the relationship to be different for marriages before 1981 compared with marriages after 1990, as reported in the second column of the table, we see that the association in arsenic levels within the couples has not changed across the marriage cohorts defined by the presence of arsenic contamination in tubewell water at the time of marriage.

We estimate (26) for households that have at least one prime-age adult, aged 18-59, of either gender. There are 720 households meeting this criterion. 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.[33] Table 12 presents descriptive statistics for the sample of 720 households for the set of variables used in the specification. Total household expenditures are the sum of household expenditures on food in the month prior to the survey multiplied by 12 plus total nonfood expenditures for the year, based on detailed questions for 149 foods and 61 nonfood items, including durable goods, school fees, housing costs.

The first column of Table13 presents the OLS estimates of the expenditure equation (26) and the second column the LIML estimates. The (marginally) statistically-preferred LIML estimates conform to the patterns implied by the division of labor in Bangladesh households. The test statistics indicate only the retained arsenic of prime age males matter for total household expenditure - the β_1 and β_2 coefficient are jointly significant at .02 level, while the coefficients β_1 and β_2 associated with prime-age women are not statistically significantly different from zero. The β_2 coefficient is negative, consistent with h_1 being negative and in accord with the findings with respect to cognitive and physical performance. 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%. For a household with male two earners, the household expenditure increase would be double that to 18%.[34]

c. Arsenic and household productivity

The estimate of β_2 indicates that men with higher retained arsenic have significantly lower market productivity. Because women do not earn, the estimates of (26) are uninformative about productivity effects for women in household production and thus the estimate of β_2 may not fully account for the harm arsenic does to household welfare, which also depends importantly on home-produced goods. In this section we test propositions 2-4 of the model by estimating the determinants of the whether or not the household chooses a non tubewell source of cooking water using the same sample of households as was used to obtain the estimates of the determinants of household expenditures. The equation we estimate is

$$W_j = \gamma_1 A_{ijm} + \gamma_2 A_{ijm} N_m + \gamma_3 A_{ijf} + \gamma_4 A_{ijf} N_f + \boldsymbol{Z}\gamma_5 + \varsigma_j, \tag{27}$$

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

[33]Of these, 236 households have both prime-age men and women of with different lineages; 92 contain only prime-age women and 131 only prime-age men.

[34]74% of households have only one prime-age male. Only 9% of households have more than two men in this age group. We tested and could not reject the multiplicative functional form imposed in (26) by estimating a more general specification allowing the β_2 coefficients to differ by number of males.

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 reduces retained arsenic (Table 3). Given these findings, proposition 2 of the model indicated that if higher arsenic retention among women was 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 stated that the sign of γ_2, the effect of male-earner retained arsenic on water source choice, indicated whether household and purchased goods were substitutes ($\gamma_2<0$) or complements ($\gamma_2>0$), given that no men contribute significantly to household production. And proposition 4 indicated that the effects of point-source water purification on household income would be underestimated by β_2, the effect of male earner retained arsenic on total household expenditures, if $\gamma_4>0$.

Because, as we have found, use of non tubewell water reduces retained arsenic, the estimates of the γ's in equation (27) are likely to be negatively biased. Comparison of the OLS and LIML estimates of (27) in Table 14 confirm this. The LIML estimates, which make use of the lineage instruments as described for the estimation of household consumption expenditures, indicate that both γ_2 and γ_4 are positive and statistically significant. The LIML point estimates for γ_2 and γ_4 are three times and 1.5 times larger, respectively, than their OLS counterparts, with the OLS γ_2 estimate not statistically significant. And indeed the correlations of each of the retained arsenic variables are significantly negatively correlated with the residuals ς_j.

The LIML γ estimates indicate that arsenic contamination significantly reduces productivity in household goods production and that household goods and purchased goods are complements. These estimates imply that, 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. However, as proposition 4 indicates, the estimates also imply that reductions in arsenic in the source of water, because they would decrease water purification effort, would increase household overall income or consumption by less than the 9%. Thus 9% is an upper bound estimate for the household income gain from the elimination of arsenic in the water sources used by rural households in Bangladesh.

9. Conclusion

Tests of the principal water sources in rural Bangladesh indicate that a substantial fraction of the Bangladesh population in rural areas is ingesting large amounts of inorganic arsenic. Most of the consequences of arsenic ingestion from drinking and consuming foods grown in and cooked with arsenic-

contaminated water, however, are hidden. Bangladeshi citizens do not know the amounts of retained arsenic they have stored in their bodies, as conventional morbidity and physical symptoms are generally unrelated to retained arsenic levels. Neither do they know their own abilities to metabolize ingested arsenic, the effects of diet on the amount of arsenic they consume, metabolize, and excrete, or the consequences for their productivity at work or in the home. The productivity costs are also hidden from researchers and policy-makers - most men in Bangladesh do not earn wages and few women participate in the labor market. Yet, almost all men earn and women contribute importantly to the production of home goods and services, including water retrieval and treatment. In this paper we have used recent evidence from the molecular genetics literature on the genetic basis of arsenic metabolism, and unique information on family links among respondents living in different environments from a large panel survey, within a theoretical framework incorporating optimizing behavior, to uncover some of the hidden costs of arsenic poisoning in Bangladesh.

Based on toenail clipping biomarkers, we found that the levels of retained arsenic in the rural Bangladesh population are nearly 20 times those measured in subjects residing in the United States, vary considerably across individuals but are no more elevated in high- than in low-wealth sub-populations. Using these data and information on individual consumption, we implemented our identification methodology by first obtaining estimates of the causal effects of the consumption of seven food groups, choice of water source and smoking on arsenic retention. These estimates show that diet matters for arsenic retention, with the consumption of grains, the staple of the Bangladesh diet, significantly increasing, but tubers, in accord with the epidemiology literature, reducing measured arsenic concentrations. We also find that using non tubewell water sources for cooking reduces retained arsenic concentrations by 18%.

We then examined the relationship between retained arsenic and measures of individual cognitive and physical capabilities, schooling attainment, health, occupational choice, entrepreneurship and income using as instruments measures of arsenic retention net of the influence of diet, water-source choices, and village fixed effects of separated family members. We found that performance on a test assessing cognition was significantly lower the higher are levels of retained arsenic and that the significant negative effects on cognition are manifested in lower schooling attainment especially for males. Young men with higher concentrations of retained arsenic were also significantly less likely to have a skilled occupation or to become entrepreneurs. We showed that these results were not due to a direct correlation in genes associated with arsenic methylation and genes associated with cognition and strength by estimating correlations among the relevant genes using newly-available genomic data on human populations. We also

could find no relationship between arsenic retention and BMI and conventional morbidity symptoms reported in our survey data, consistent with the productivity effects of arsenic contamination being hidden in the population.

Using information on household expenditures, we obtained estimates of the effects of variation in retained arsenic among prime-age males regardless of their status as wage or self-employed workers. The estimates indicated that lowering the amount of retained arsenic among Bangladesh prime-age males to those levels in uncontaminated countries would increase total household expenditures by 9% for each prime age male. We also found that retained arsenic reduces the productivity of women in home production, and that home goods and purchased goods are complements. These latter results imply that the estimated increase in total household expenditures obtained by eliminating arsenic, which form only a part of total household consumption, closely corresponds to the overall consumption benefits to the household.

What do our estimates imply for policies that may reduce the economic burden of arsenic-contaminated wells? First, while we have found that diet matters for arsenic retention, the beneficial effects of dietary changes are relatively small - for example, doubling tuber consumption would only reduce retained arsenic by 5%. Second, while we obtained estimates that implied that if everyone switched from tubewells as their source of cooking water, retained arsenic would decline by 18%, 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, and would only increase incomes by less than 2%.

The convenient provision of non-contaminated water that would eliminate the problem of arsenic contamination, through a one-time investment in the provision of piped water or simple treatments of contaminated water, is obviously costly in a rural population. Our estimates suggest, however, that the economic benefits are not trivial, even ignoring all of the potential health costs in later life (e.g., cancers). The annual benefits include a rise in earnings for every male worker of $54 per year. If the value of time of women is half of that of men, and the loss in productivity is similar, recalling that we found evidence of a significant loss in nonmarket productivity for women, that would imply an additional benefit of $27 per year. Added to this would be a reduction in women's time spent fetching water, which our data suggests is on average 20 minutes per day. Valued at half of the male wage, the freeing up of women's time from having convenient, clean water available would add an additional benefit of $13 per year. The present discounted value of the per-household gain of $94 over twenty years ranges from $1400 to $1000 for discount rates of from 3% up to 8%. The added consumption through productivity improvements, of

course, assumes that there would be no economic growth. As income grows, the hidden cost of not improving clean water availability also grows. And the loss in human capital from the reductions in cognition and human capital investment suggest that growth rates may also be reduced.

References

Agusa, T, Fujihara J, Kunito T, Takeshita T, Iwata H. 2011. Individual Variations in Inorganic Arsenic Metabolism Associated with AS3MT Genetic Polymorphisms. Int. J. Mol. Sci. 12, 2351-2382.

Ahsan H, Chen Y, Kibriya MG, Slavkovich V, Parvez F, Jasmine F, Gamble MV, Graziano JH. 2007. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh.Cancer Epidemiol Biomarkers Prev. 16(6):1270-8.

Argos, M. 2011. Genetic Susceptibility to Arsenic Exposure and Arsenical Skin Lesion Prevalence in Bangladesh. PhD thesis, Columbia University.

Asadullaha, M. N. and N. 2011. Poisoning the mind: Arsenic contamination of drinking water wells and children's educational achievement in rural Bangladesh. Economics of Education Review 30(2011): 873-888.

Bhattacharya S, Gupta K, Debnath S, Ghosh UC, Chattopadhy D & Mukhopadhyay A. 2012. Arsenic bioaccumulation in rice and edible plants and subsequent transmission through food chain in Bengal basin: a review of the perspectives for environmental health. Toxicological & Environmental Chemistry 94(3):429-441.

Brammer H, Ravenscroft P. 2009. Arsenic in groundwater: a threat to sustainable agriculture in South and South-east Asia. Environ Int. 35(3):647-54.

British Geological Survey. 1999. Groundwater Studies for Arsenic Contamination in Bangladesh Phase 1 British Geological Survey [http://www.bgs.ac.uk/arsenic/home.html].

Chanda S, Ganguli B, Dasgupta UB, and Mazumder DG. 2011. Glutathione–S-transferase polymorphic status modifies the arsenic induced clinical manifestation in Nadia District of West Bengal, India. Journal of Public Health and Epidemiology, 3(10):424-428.

Chen Y, Graziano JH, Parvez F, Hussain I, Momotaj H, van Geen A, Howe GR, Ahsan H. 2006. Modification of risk of arsenic-induced skin lesions by sunlight exposure, smoking, and occupational exposures in Bangladesh. Epidemiology. 17(4):459-67.

Chung JS, Kalman DA, Moore LE, Kosnett MJ, Arroyo AP, Beeris M, Mazumder DN, Hernandez AL, Smith AH. 2002. Family correlations of arsenic methylation patterns in children and parents exposed to high concentrations of arsenic in drinking water. Environ Health Perspect. 110(7):729-33.

Engström S K, Broberg K, Concha G, Nermell B, Warholm M, Vahter M. Genetic polymorphisms influencing arsenic metabolism: Evidence from Argentina. Environ. Health Perspect. 2007, 115, 599–605.

Farid ATM, Roy KC, Hossain KM, and Sen R. 2003. A Study of Arsenic Contaminated Irrigation Water and its Carried Over Effect on Vegetable. In Ahmed MF, Ali MA, and Adeel Z (eds) Proceedings of the International Symposium on Fate of Arsenic in the Environment. Bangladesh University of Engineering and Technology (BUET) and United Nations University (BUET, Dhaka, Bangladesh, 2003), 113-121.

Field, E., R Hussam, and R. Glennerster. 2010. Throwing the Baby out with the Drinking Water: Unintended Consequences of a National Arsenic Mitigation Campaign in Bangladesh. December 2010. Working paper, Harvard University.

González-Neira A, Ke X, Lao O, Calafell F, Navarro A, Comas D, Cann H, Bumpstead S, Ghori J, Hunt S, Deloukas P, Dunham I, Cardon LR, Bertranpetit J. 2006. The portability of tagSNPs across populations: a worldwide survey. Genome Res. 16(3):323-30.

Heck JE, Gamble MV, Chen Y, Graziano JH, Slavkovich V, Parvez F, Baron JA, Howe GR, Ahsan H. 2007. Consumption of folate-related nutrients and metabolism of arsenic in Bangladesh. American Journal of Clinical Nutrition 85(5):1367-74

Hopenhayn-Rich C, Biggs ML, Smith AH, Kalman DA, Moore LE. 1996. Methylation study of a population environmentally exposed to arsenic in drinking water. Environ Health Perspect. 104(6):620-8.

Hossain M.F. 2006. Arsenic contamination in Bangladesh-An overview. Agriculture, Ecosystems and Environment 113(1-4): 1-16.

Huq SMI , Joardar JC, Parvin S, Correll R, and Naidu R. 2006. Arsenic Contamination in Food-chain: Transfer of Arsenic into Food Materials through Groundwater Irrigation. J Health Popul Nutr 24(3):305-316.

Jackson BP, Taylor VF, Karagas MR, Punshon T, Cottingham KL. 2012. Arsenic, organic foods, and brown rice syrup. Environ Health Perspect. 120(5):623-6.

Kelley, K.A., Plank, T., Ludden, J.N., Staudigel, H., 2003. Composition of altered oceanic crust at ODP Sites 801 and 1149. Geochem. Geophys. Geosys. 4, doi:10.1029/2002GC000435.

Kile ML, Houseman EA, Rodrigues E, Smith TJ, Quamruzzaman Q, Rahman M, Mahiuddin G, Su L, Christiani DC. 2005. Toenail arsenic concentrations, GSTT1 gene polymorphisms, and arsenic exposure from drinking water. Cancer Epidemiol Biomarkers Prev.;14(10):2419-26.

Kile ML, Houseman EA, Breton CV, Smith T, Quamruzzaman Q, Rahman M, Mahiuddin G, Christiani DC. (2007). Dietary arsenic exposure in Bangladesh. Environmental Health Perspectives 115(6):889-93.

Kile ML, Hoffman E, Hsueh YM, Afroz S, Quamruzzaman Q, Rahman M, Mahiuddin G, Ryan L, Christiani DC. 2009. Variability in biomarkers of arsenic exposure and metabolism in adults over time. Environ Health Perspect. 117(3):455-60.

Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. Trends in Genetics, 27(8): 323-31.

McCarty KM, Chen YC, Quamruzzaman Q, Rahman M, Mahiuddin G, Hsueh YM, Su L, Smith T, Ryan L, Christiani DC. 2007. Arsenic methylation, GSTT1, GSTM1, GSTP1 polymorphisms, and skin lesions. Environ Health Perspect. 115(3):341-5.

Pierce BL, Argos M, Chen Y, Melkonian S, Parvez F, Islam T, Ahmed A, Hasan R, Rathouz PJ, Ahsan H. 2011. Arsenic exposure, dietary patterns, and skin lesion risk in bangladesh: a prospective study. Am J Epidemiol. 173(3):345-54.

Pitt, M. M., M. R. Rosenzweig and N. Hassan. forthcoming. Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy. American Economic Review.

Rodrigues EG, Kile M, Hoffman E, Quamruzzaman Q, Rahman M, Mahiuddin G, Hsueh Y, Christiani DC. 2012. GSTO and AS3MT genetic polymorphisms and differences in urinary arsenic concentrations among residents in Bangladesh. Biomarkers 17(3):240-7.

Sidak, Z. 1967. Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association 62(318):626-633.

Smith AH, Lingas EO, Rahman M. 2000. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. Bulletin of the World Health Organization 78(9):1093-103

Steinmaus C, Moore LE, Shipp M, Kalman D, Rey OA, Biggs ML, Hopenhayn C, Bates MN, Zheng S, Wiencke JK, Smith AH. 2007. Genetic polymorphisms in MTHFR 677 and 1298, GSTM1 and T1, and metabolism of arsenic. J Toxicol Environ Health A. 70(2):159-70.

Stock, J. H., and Yogo, M. 2001, Testing for Weak Instruments in Linear IV Regression, unpublished manuscript, Harvard University.

Stock J. H., J.H. Wright and M. Yogo. 2002 A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. Journal of Business and Economic Statistics 20(4): 518-529.

Vahter M. (2000) Genetic polymorphism in the biotransformation of inorganic arsenic and its role in toxicity. Toxicology Letters 112-113:209-17.

Xing J, Witherspoon DJ, Watkins WS, Zhang Y, Tolpinrud W, Jorde LB. 2008. HapMap tagSNP transferability in multiple populations: general guidelines. Genomics. 2008 92(1):41-51.

Zablotska LB, Chen Y, Graziano JH, Parvez F, van Geen A, Howe GR, Ahsan H. 2008. Protective effects of B vitamins and antioxidants on the risk of arsenic-related skin lesions in Bangladesh. Environmental Health Perspectives 116(8):1056-62

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

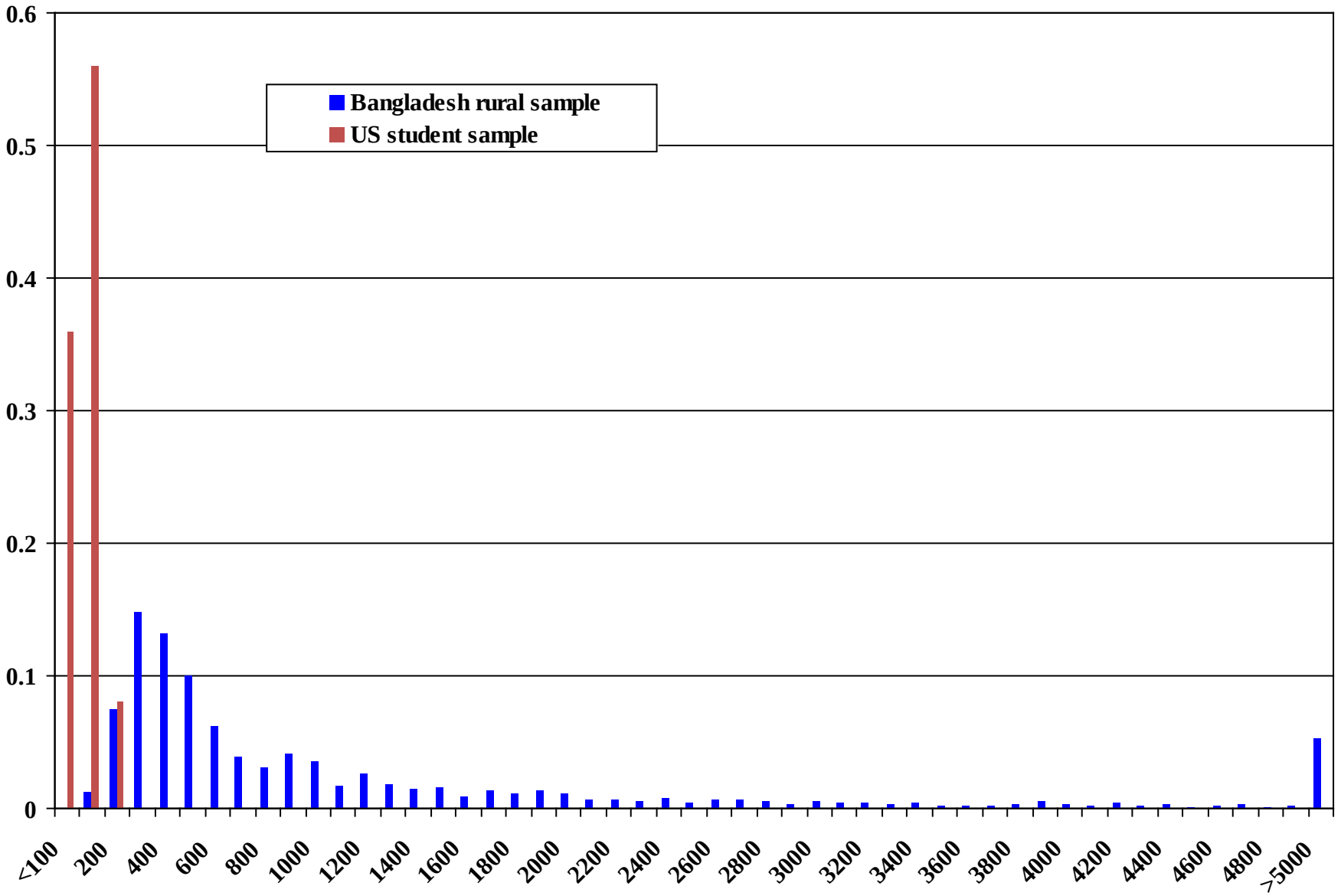

Figure 2. Mean *As* Concentrations (ppb) by Owned Landholdings (square decimeters)

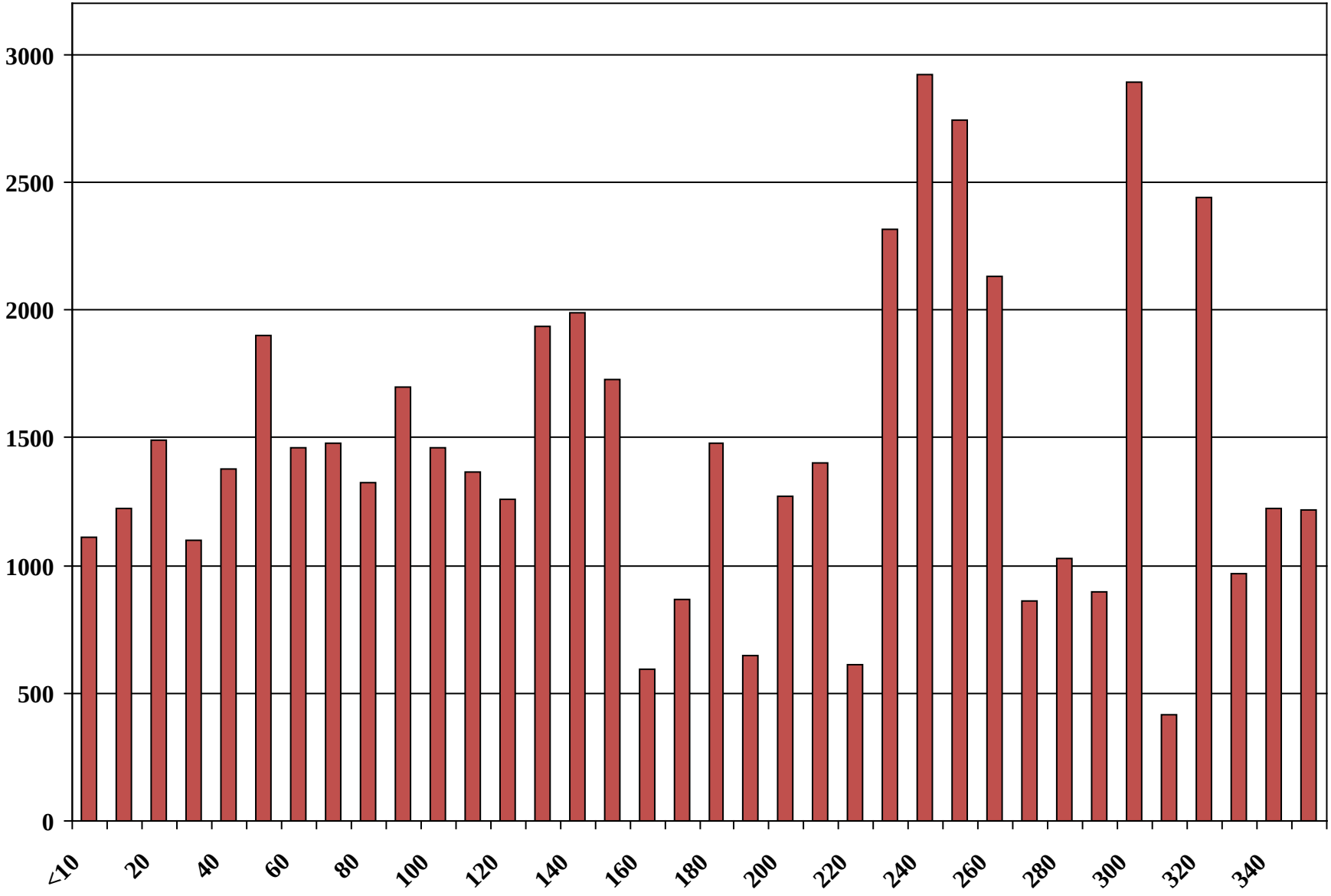

Figure 3. Locally-weighted Estimates of the Effects of Non-Coresident Lineage *As* Endowments on Respondent *As* Concentrations, by Lineage *As* Endowment Size

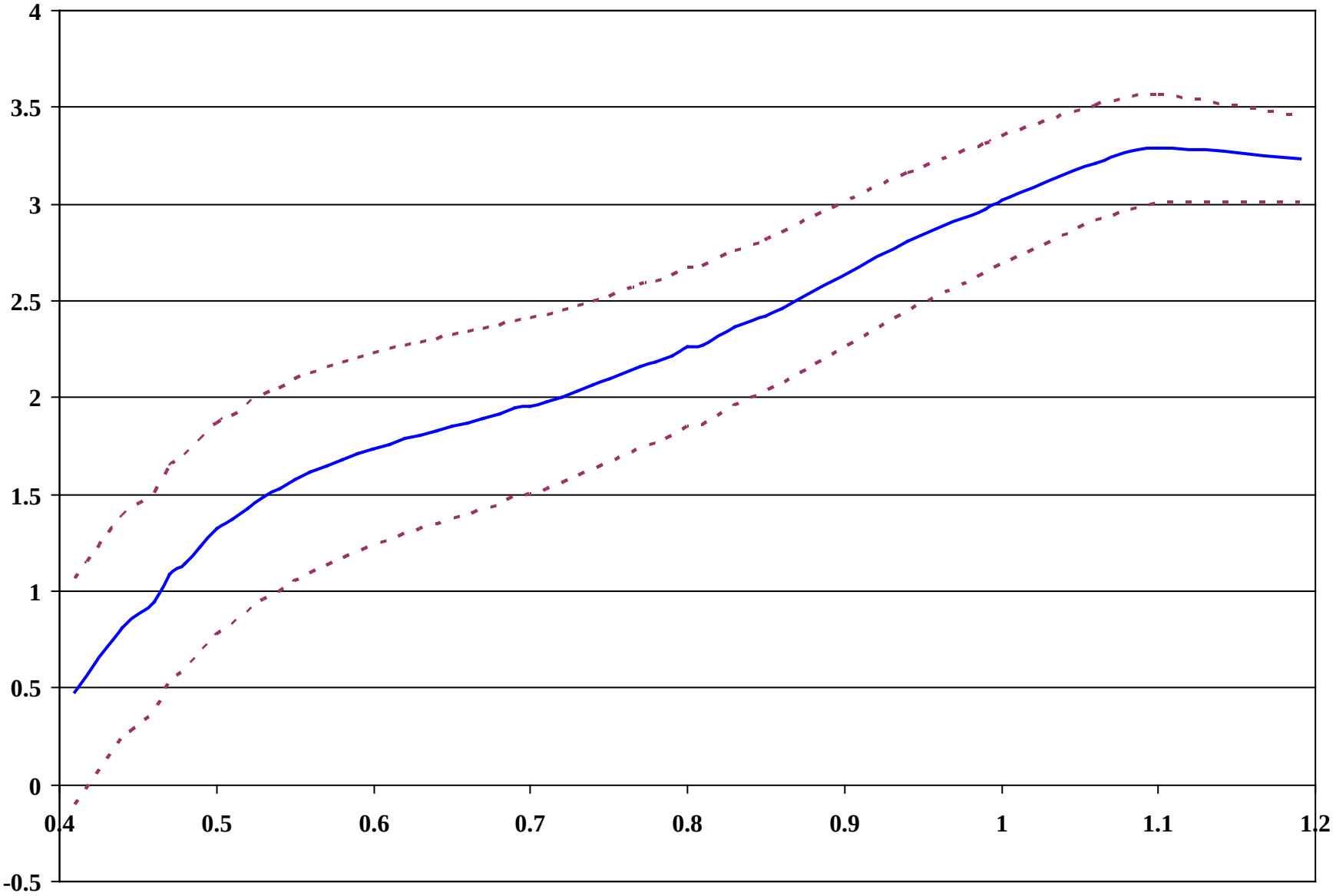

Figure 4. Locally-weighted IV Estimates of the Effects of *As* on Years of Schooling Completed for Men, by Age (Cohort)

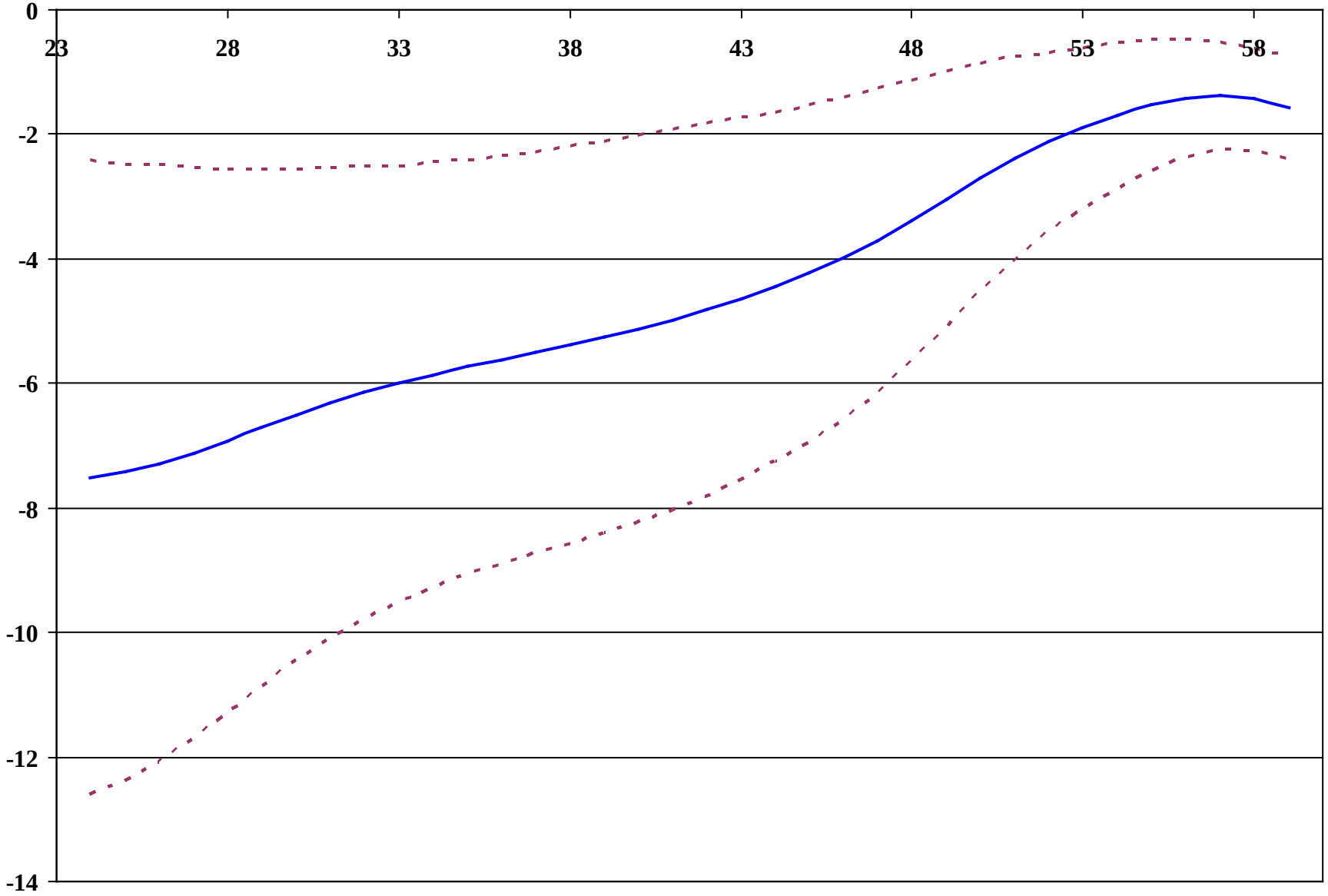

Figure A. Height (cm) by Year Attained Age 22 and Gender, 1970 – 2004 for Respondents Aged 22-59 in 2007-8 (Lowess-smoothed)

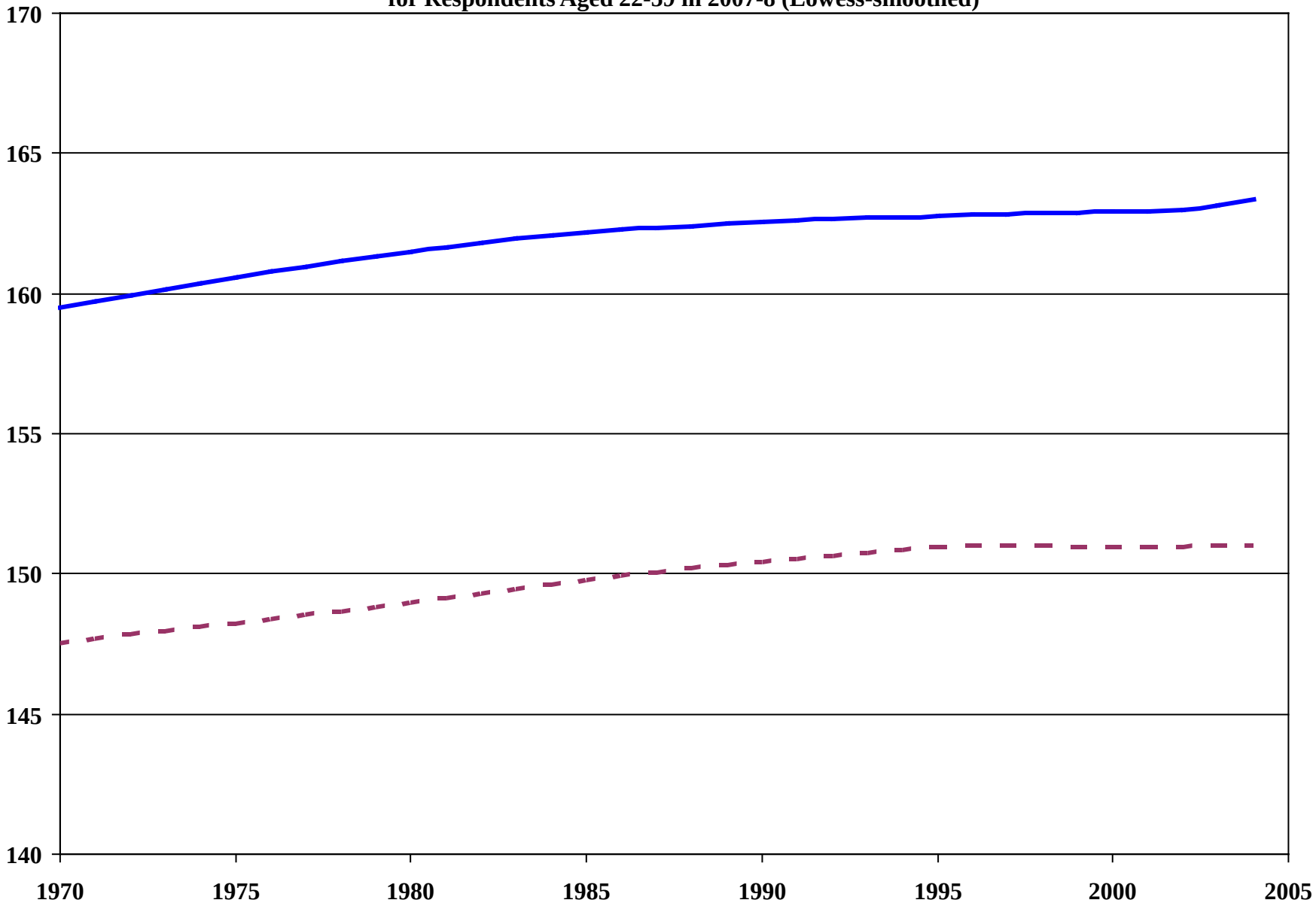

Figure B. BMI in 1981-82 and 2007-8, by Gender

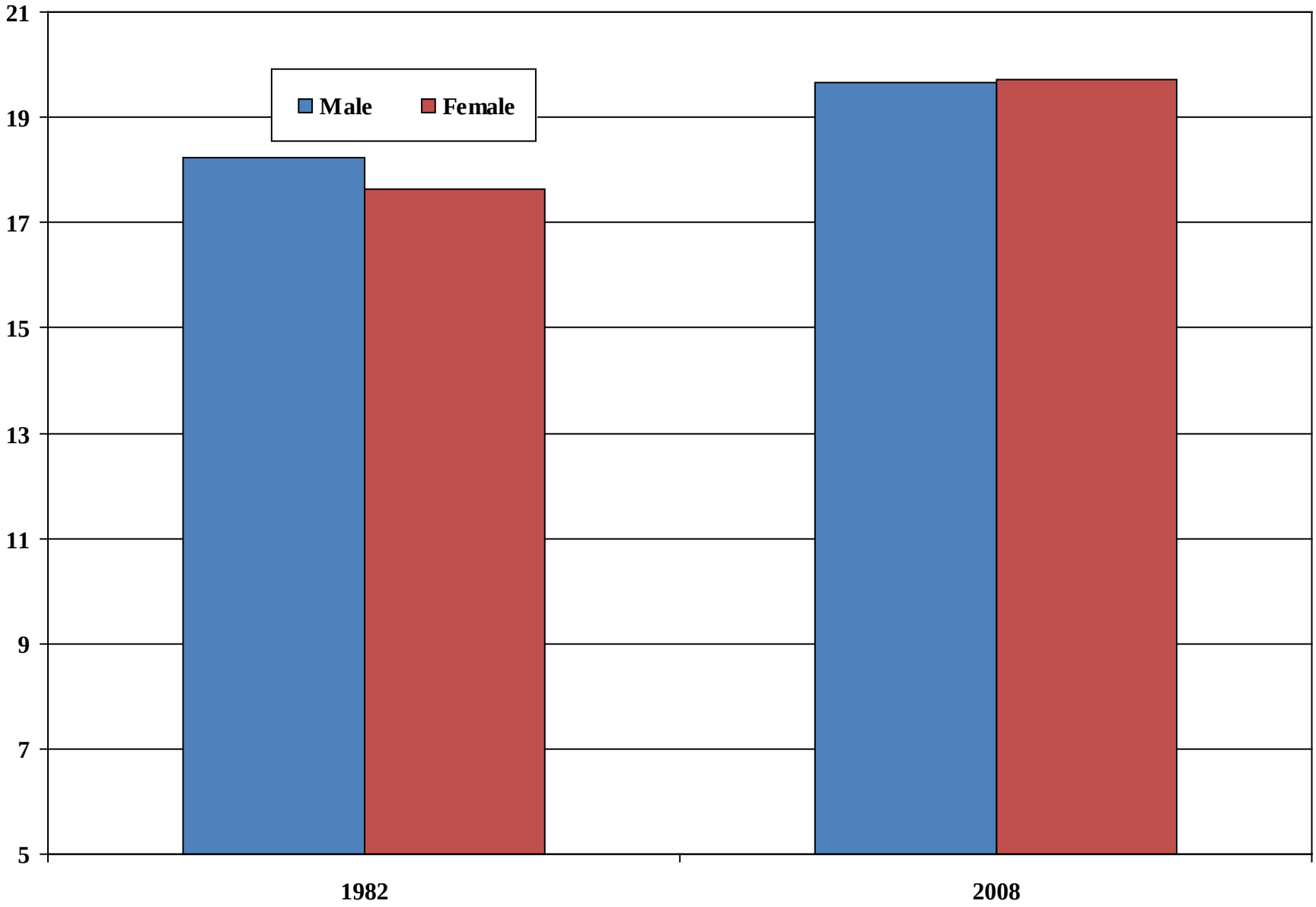

Table 1: Respondent Characteristics (Means and Standard Deviations): Men and Women Aged 18-59

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

Standard deviation in parentheses.

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.

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 (2.81)	-6.93 (1.66)
Mean log non co-resident lineage (NCL) *As* residuals squared	-	5.44 (2.25)
Age	.0219 (1.27)	.0189 (1.05)
Age squared	-.0287 (1.19)	-.0250 (1.01)
Value of owned landholdings ($x10^{-7}$)	.854 (2.02)	.725 (2.00)
Male	.0291 (0.34)	.0382 (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]

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.

Table 4
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 (3.32)	-.669 (6.36)	-.132 (1.65)
Age	-.133 (4.76)	-.122 (3.91)	-.134 (4.86)
Age squared	.109 (3.08)	.0935 (2.27)	.110 (3.17)
Value of owned landholdings ($x10^{-7}$)	3.55 (4.61)	3.97 (4.73)	3.51 (4.60)
Male	.904 (8.89)	.914 (7.74)	.900 (8.91)
N	1519	1519	1519
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]

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. NCL=non-coresident lineage.

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 17-59. Absolute values of *t*-ratios in parentheses clustered at the village level.

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

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 17-59.
Absolute values of *t*-ratios in parentheses clustered at the village level.

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

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 [.0168]	1.39 [.243]	1.25 [.168]	4.98 [.0337]
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 17-59.
Absolute values of t-ratios in parentheses clustered at the village level.

Table 8: Tests of Linkage Disequilibrium between SNPs related to arsenic metabolism and SNPs related to IQ, body mass (BMI) and height, and muscle development and 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	
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
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
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)

(http://hapmap.ncbi.nlm.nih.gov/index.html.en). Computed with PLINK software from the Broad Institute.

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.

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

Gender	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 ($x10^{-7}$)	5.20	4.99	-2.80	-.395	76.5	48.0
	(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 17-59.
Absolute values of *t*-ratios in parentheses clustered at the village level.

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.

Table 12
Household Characteristics

Variable	Mean
Total annual household expenditures (*Tk*)	54,290 (45,015)
Cooking water source not a well	.235 (.424)
Value of landholdings (*Tk* x 10^{-4})	37.2 (78.4)
Number of males aged	1.72 (1.10)
Number of females aged	1.52 (.836)
Number of males 18-59	.672 (.755)
Number of females 18-59	.692 (.739)
Average age of household males	17.8 (6.09)
Average age of household females	15.6 (7.82)
N	720

Standard deviation inparentheses.

Table 13
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 (0.85)	.00113 (0.17)
Mean Log *As* of Men Aged 18-59 x Number of Men Aged 18-59 (m*As* x mp)		.0279 (1.55)	-.0941 (2.65)
Mean Log *As* of Women Aged 18-59 (f*As*)		.0481 (1.48)	.00109 (0.14)
Mean Log *As* of Women Aged 18-59 x Number of Women Aged 18-59 (f*As* x fp)		-.0300 (1.65)	-.00108 (0.05)
Number of Men Aged 18-59 (mp)		-.107 (0.85)	.673 (2.77)
Number of Women Aged 18-59 (fp)		.233 (1.75)	-.00393 (0.03)
Value of owned landholdings ($x10^{-7}$)		1.54 (6.82)	1.91 (9.73)
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)
N		720	720
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]

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

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

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.

Appendix A

Analytical Methods for the Determination of *As* in Toenail Solutions

Toenail samples were cleaned and digested by Robert Correra using the Environmental Chemistry Facility at Brown University under the direction of David Murray. The nail clippings were prepared for As analysis by ICPMS using a cleaning procedure modified from one used at the Trace Element Analysis Laboratory at Dartmouth College (http://www.dartmouth.edu/~toxmetal/assets/pdf/teamethods.pdf). Toenail clippings from an individual, targeting a combined weight between 10-50 mg, were placed in a 7ml polyethylene vial (Fisher #03-337-20) after removing any visible dirt. 2 ml of Optima grade acetone was added to each sample and vials were placed in a rack in an ultrasonic bath for 20 min. The acetone was removed with a vacuum siphon and 2 ml 1% solution of Triton X-100 was added. Vials were sonicated again for 20 minutes, 5 ml of deionized (DI) water was added, and the supernatant was removed with a vacuum siphon. Samples were rinsed 4 more times with 5 ml DI water, decanting the water after shaking clippings and water. Rinsed samples were dried in a convection oven at 60°C for 12 hrs.

The cleaned nail sample was weighed (0.1 mg accuracy), placed in preweighed and labeled 15 ml polypropylene centrifuge tube (VWR #21008-103), and 1.0 ml Optima grade Nitric Acid was added. Racks containing batches of 100 tubes were placed in an 80°C water bath for 120 min. After cooling for 30 min, 0.2 ml of optima grade hydrogen peroxide was added, and tubes were placed in the 80°C bath for 30 minutes. After cooling, 6 ml of DI water was added to each tube and the final weight was recorded. Batches of 100 tubes include 91 toenail samples, 2 samples each of National Research Council Canada reference material DORM-3 and DOLT-4, one set of 0.5, 1, and 2 ppm As standards, and 2 blanks.

Solutions were analyzed for As concentration using a Thermo X-Series II quadrupole inductively-coupled plasma mass spectrometer (ICP-MS) equipped with collision cell technology (CCT) at the Graduate School of Oceanography, University of Rhode Island, under the supervision of Professor Kathleen Kelley. The CCT was flushed with a He-H gas mix to break up the 40Ar35Cl interference on 75As. Daily assessment of the efficiency of the collision cell, monitoring mass 75 in an ultra-pure 2% HCl solution , showed that the CCT reduced the interference to <0.02% 40Ar35Cl/35Cl. In practical terms, this translated to count rates of <1 count per second attributed to Cl in the unknown solutions, effectively removing the interference. All samples were run using identical analysis and data reduction protocols, similar to procedures described by Kelley et al. (2003). A procedural blank was subtracted from all raw data, and count rates were corrected for instrumental drift using an external drift correcting solution analyzed every 5 samples, which was matrix-matched to the unknown samples and spiked with 15 ppb As. Concentrations of As in unknown solutions were determined by calibrating against four single-element As solution standards spanning a range from 0.5 to 100 ppb. Calibration curves were linear, with r>0.9990 for all analytical runs. Arsenic concentrations in unknown solutions were then corrected for the dilution of each starting sample weight, to yield the concentration of As in the raw toenail material. Each unknown solution was analyzed twice in every analytical session, and within-session reproducibility was <10% rsd for these replicate analyses (on average, within 5% rsd). Some samples contained enough material to perform replicate dissolutions of the starting material (n=184), and these replicate dissolutions are in excellent agreement, on average within 15% rsd. Lower precision for replicate dissolutions of the raw material is not surprising, likely resulting from heterogeneities within the toenails, which were dissolved as whole clippings without attempting to homogenize the material in advance.

Appendix Table A
SNPs related to arsenic methylation

SNP	Gene	Chromosome	Gene Function	Relation to other health conditions
rs11191439 rs7085104	AS3MT	10	Arsenite methyltransferase , a critical enzyme for arsenite biotransformation, is encoded by the AS3MT gene	none
rs1801133 rs1801131	MTHFR (methylenetetrahydrofolate reductase)	1	Provides instructions for making the enzyme methylenetetrahydrofolate reductase . This enzyme is important for a chemical reaction involving forms of folate (folic acid) which has been linked to both arsenic metabolism and toxicity.	Neural tube defects (NTD), spina bifida, and homocystinuria, all relatively rare inherited disorders.
rs4925 rs11509438	GSTO1	10	Encodes glutathione S-transferase omega-1 enzyme. This enzyme reduces susceptibility to environmental toxins such as arsenic and some drugs.	Mutations linked with an increase in a number of cancers, likely due to an increased susceptibility to environmental toxins and carcinogens.

Appendix Table B

SNP Locations

SNP type	SNP number	Chromosome	Base-pair position
IQ			
	rs363039	20	10168496
	rs4680	22	18331271
	rs2760118	6	24611569
	rs821616	1	230211221
	rs1018381	6	15765049
Body mass & height			
	rs724016	3	142588260
	rs143384	20	33489170
	rs1351394	12	64638093
	rs7689420	4	145787802
	rs6449353	4	17642586
	rs1421085	16	52358455
	rs211683	1	75834515
	rs988712	11	27519958
Muscle development and physical strength			
	rs7843014	8	141859663
	rs7460	8	141738042
	rs1800169	11	58148077
	rs3808871	9	34581425
	rs1815739	11	66084671
Arsenic			
	rs11191439	10	104628713
	rs7085104	10	104618863
	rs4925	10	106012779
	rs11509438	10	106017049
	rs1801133	1	11777063
	rs1801131	1	11778965
Source: SNP location based on query at http://www.broadinstitute.org/mpg/snap/ldsearchpw.php			

Appendix B

Single-Nucleotide Polymorphisms (SNPs) identified for the linkage disequilibrium analysis

A. General Intelligence

Rs363039: The SNAP-25 (synaptosomal-associated protein of 25 kDa) gene plays an integral role in synaptic transmission and is highly expressed by neurons in the hippocampus. The hippocampus is a critical part of the central nervous system associated with learning and memory processes. Damage to the hippocampus selectively impairs the ability to learn and remember. SNAP25 encodes a presynaptic protein with a role in regulation of neurotransmitter release. In addition, the SNAP-25 gene lies in a linkage area implicated in human intelligence. Mouse mutants with deficient expression of SNAP-25 demonstrate impairment of neuronal development, short-term synaptic plasticity, and cognitive function (Johansson *et al.* 2008)). Recent studies have suggested a possible involvement of SNAP-25 in learning and memory, both of which are key components of human intelligence. In Gosso *et al.* (2006), two independent family-based Dutch samples of 391 (mean age 12.4 years) and 276 (mean age 37.3 years) subjects, respectively, were genotyped for SNPs in the SNAP-25 gene. Using a family-based association test, a strong association was found between rs363039 and two other SNPs in the SNAP-25 gene and intelligence, measured by standard intelligence measures, for both sexes. Rs363039 and another SNP (rs363050) were associated with IQ in both the young cohort and the independent adult cohort, showing association in the same direction and the same order of magnitude. Söderqvist *et al.* (2010) found that rs363039 was associated with superior working memory, as measured by a battery of tests, higher performance on fluid intelligence, as measured by Raven's Progessive Matrices, an abridged version of which is used in this paper, and on gray matter density, as measured by MRI scans. Cagliani *et al.* (2011) analysed rs363039 variants in two Italian cohorts in different age ranges and observed a significant genotype effect for rs363039 on verbal performance in females alone. Beunders *et al.* (2012) find a significant association of this SNP with the extremes of intelligence, and conclude that SNAP-25 variants affect both variation in normal intelligence and mental retardation as well as high IQ.

Rs4680: This is the only common functional SNP on the human COMT (catechol-O-methyltransferase) gene. The COMT gene has been one of the most extensively studied candidate genes in relation to individual differences in cognitive ability. The Valine158Methionine (Val158Met) polymorphism of the COMT gene results in lower enzymatic activity and higher dopamine availability in those with the Met allele (rs4680(A)) as compared to the Val allele (rs4680(G)) (Dumontheil *et al.*, 2011). The Met allele is also associated with better performance and reduced prefrontal cortex activation during working memory tasks in adults, as measured by a battery of tests, and higher IQ. Consistent with previous evidence, Dumontheil *et al.*, 2011 also find that this polymorphism affects both adolescent and adult gray matter volumes as measured by MRI scans. Mitaki *et al.* (2012) find that the rs4680 SNP affects the variability in executive function, as assessed by the FAB score, in a sample of 964 individuals. Executive function is the theorized cognitive system that controls and manages other cognitive processes and encompasses planning, working memory, attention, problem solving, verbal reasoning, inhibition, mental flexibility, multi-tasking, initiation and monitoring of actions. These executive functions are located primarily in the prefrontal regions of the frontal lobe. The FAB score of executive function is derived from six subtests exploring conceptualization, mental flexibility, motor programming, sensitivity to interference, inhibitory control, and environmental autonomy. In addition, the

tendency to choose immediate over delayed rewards (*delay-discounting*) has been associated with rs4680 in a number of papers. Very recently, Smith and Boettiger (2012) find that age modulates the effects of rs4680 on delay-discounting behavior. In particular, among those with the Met allele, delay-discounting was negatively correlated with age from late adolescence to adulthood, while among Val allele individuals delay-discounting was positively correlated with age across this range.

Rs2760118: A functional polymorphism in the SSADH(succinate-semialdehyde dehydrogenase gene, also known as ALDH5A1 (aldehyde dehydrogenase 5 family, member A1). Succinate-semialdehyde dehydrogenase (SSADH) deficiency is a rare cause of learning disability, and has been related to cognitive ability more generally. Recently, Rizzi *et al.* (2011) found an association between this SNP and IQ as measured by testing in a sample of 4963 subjects. Plomin *et al.* (2004) found that the minor allele of this SNP was significantly less frequent in high-IQ cases. They note that this allele encodes a lower activity enzyme than the major allele and conclude that higher SSADH activity is associated with higher intelligence across the general population.

Rs821616: DISC1 (Disrupted in schizophrenia 1) gene variants have been linked to altered cognition, brain structure and brain function. A new study (Singh *et al.*, 2011), experimenting with multiple model systems using mouse-, zebra fish- and human cells, shows how DISC1 variants impair signaling pathways and disrupt brain development. Evidence has emerged from genetic studies showing a relationship between DISC1 and quantitative measures of working memory, cognitive aging, gray matter volume in the prefrontal cortex, and abnormalities in hippocampal structures and function. For example, Carless *et al.* (2011) found that DISC1 polymorphisms were associated with cortical gray matter thickness (as measured by MRI) within the parietal, temporal and frontal lobes, and measures of working memory.

Rs1018381: The DTNBP1 (human dystrobrevin-binding protein 1) gene contains the coding region for the dysbindin protein. Because DTNBP1 is widely expressed in the brain, it is believed to play a role in a number of brain functions. Animal studies have demonstrated that dysbindin plays a critical modulatory role in synaptic transmission. Zhang *et al.* (2010) perform a meta-analysis of genetic variation in DTNBP1 and general cognitive ability. Pooled effect estimates suggest that the minor allele carriers of this SNP had lower cognitive ability scores than the major allele carriers. In Luciano *et al.* (2009) the association between DTNBP1 genotype and cognitive abilities was investigated in three population samples (1054 Scottish, 1806 Australian and 745 English) of varying age. There was evidence in each of the cohorts for association these associations including measures of memory and of executive function.

B. Body Mass

rs1421085: Variants in the FTO (fat mass and obesity associated) gene has been widely found to be associated with body mass index (BMI). It's presumed mechanism is through the regulation of lipid storage. The analysis of the FTO gene in mice showed that allele variation altered the energy expended for physical activity. We have chosen the rs1421085 SNP from this gene. Dina *et al.* (2007) and many others have found strong associations of BMI and variation in this SNP, and Chauhan *et al.* (2011) have confirmed these findings for an Indian population.

rs211683: The (KCNMA1) gene that encodes potassium channel, calcium activated, large conductance, subfamily M, alpha member, has been associated with BMI in numerous studies, as reported in the meta-analysis of Jiao *et al.* (2011) and in other studies.

rs988712: The BDNF gene is believed to act primarily on the central regulation of food intake and has been associated with BMI in the meta-analysis of Jiao *et al.* (2011) and other studies.

C. Height

rs724016, *rs143384*, *rs1351394*, *rs7689420*, and *rs64449353*: The five height SNPs are from the large meta-analysis of Longo *et al.* (2010). In that study, a meta-analysis of genome-wide association data from 46 studies was used to identify common genetic variation associated with adult height. They then tested the most strongly associated SNPs with an additional sample of 50,074 individuals. The five SNPs used here are those with the highest p-values and consistent signs in the combined sample of 133,653 individuals. The p-values ranged from 4.5E-52 to 1.3E-27.

D. Muscle development and physical strength

rs7843014 and *rs7460*: The protein tyrosine 1 kinase-2 (PTK2) gene encodes focal adhesion kinase, a structural protein involved in lateral transmission of muscle fiber force. These two SNPs on this gene were found to be associated with variability in the strength responses to resistance training by Erskine *et al.* (2012).

rs4646994: The human Angiotensin I-Converting Enzyme (ACE) gene has been associated with many heritable traits, including physical performance. The ACTN3 gene encodes a protein that anchors actin, a muscle filament, within the muscle's framework, thus influencing muscle function (Kostek *et al.*, 2011). According to Williams et al. (2000), this may be related to an increase in slow-twitch rather than fast-twitch muscle fibers, which are more efficient in slow contraction. The lower ACE enzyme activity associated with one ACE genotype may also raise local concentrations of nitric oxide, which in turn may improve contractile function in both cardiac and skeletal muscle. Moran *et al.* (2006) shows that there is a strong association the ACE polymorphism and both handgrip strength and vertical jump.

rs1800160 and *rs3808871*: Ciliary neurotrophic factor (CNTF) is important for neuronal and muscle development, and genetic variation in the CNTF gene has been associated with muscle strength via its influence on motor unit size and firing patterns. These SNPs have been associated with muscle development and strength in Kostek *et al.* (2011) and others.

Appendix References

Beunders G., Rizzu P, Sistermans EA, Twisk JWR, van Mechelen W, Deijen JB, Heutink P, Meijers-Heijboer H, Posthuma D. 2009. Association between the SNAP25 gene and the extremes in intellectual performance; mental retardation and high IQ. Presented at the American Society of Human Genetics Meetings. Singapore.

Cagliani R, Riva S, Marino C, Fumagalli M, D'Angelo MG, Riva V, Comi GP, Pozzoli U, Forni D, Cáceres M, Bresolin N, Clerici M, Sironi M. 2012. Variants in SNAP25 are targets of natural selection and influence verbal performances in women. Cell Mol Life Sci. 69(10):1705-15

Carless MA, Glahn DC, Johnson MP, Curran JE, Bozaoglu K, Dyer TD, Winkler AM, Cole SA, Almasy L, MacCluer JW, Duggirala R, Moses EK, Göring HH, Blangero J. 2011. Impact of DISC1 variation on neuroanatomical and neurocognitive phenotypes. Mol Psychiatry. 16(11):1096-104, 1063.

Chauhan, G, Tabassum R, Mahajan A, Dwivedi OP, Mahendran Y, Kaur I, Nigam S, Dubey H, Varma B, Madhu SV, Mathur SK, Ghosh S, Tandon N, Bharadwaj D. 2011. Common variants of FTO and the risk of obesity and type 2 diabetes in Indians. Journal of Human Genetics 56, 720–726.

Conwit RA, Ling S, Roth S, Stashuk D, Hurley B, Ferrell R, Metter EJ. 2005. The relationship between ciliary neurotrophic factor (CNTF) genotype and motor unit physiology: preliminary studies. BMC Physiology 5:15.

Dina C, Meyre D, Gallina S, Durand E, Körner A, Jacobson P, Carlsson LM, Kiess W, Vatin V, Lecoeur C, Delplanque J, Vaillant E, Pattou F, Ruiz J, Weill J, Levy-Marchal C, Horber F, Potoczna N, Hercberg S, Le Stunff C, Bougnères P, Kovacs P, Marre M, Balkau B, Cauchi S, Chèvre JC, Froguel P. 2007. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet. 39(6):724-6.

Dumontheil I, Roggeman C, Ziermans T, Peyrard-Janvid M, Matsson H, Kere J, Klingberg T. 2011. Influence of the COMT genotype on working memory and brain activity changes during development. Biol Psychiatry. 70(3):222-9.

Erskine RM, Williams AG, Jones DA, Stewart CE, Degens H. 2012. Do PTK2 gene polymorphisms contribute to the inter-individual variability in muscle strength and the response to resistance training? A preliminary report. J Appl Physiol.112(8):1329-34.

Erskine, Robert M., Alun G. Williams, David A. Jones, Claire E. Stewart and Hans Degens.

Gosso MF, van Belzen M, de Geus EJ, Polderman JC, Heutink P, Boomsma DI, Posthuma D. 2006. Association between the CHRM2 gene and intelligence in a sample of 304 Dutch families. Genes Brain Behav. 5(8):577-84.

Jiao H, Arner P, Hoffstedt J, Brodin D, Dubern B, Czernichow S, van't Hooft F, Axelsson T, Pedersen O, Hansen T, Sørensen TI, Hebebrand J, Kere J, Dahlman-Wright K, Hamsten A, Clement K, Dahlman I. 2011. Genome wide association study identifies KCNMA1 contributing to human obesity. BMC Medical Genomics, 4:51

Johansson JU, Ericsson J, Janson J, Beraki S, Stanić D, Mandic SA, Wikström MA, Hökfelt T, Ogren SO, Rozell B, Berggren PO, Bark C. 2008. An ancient duplication of exon 5 in the Snap25 gene is required for complex neuronal development/function. PLoS Genet. 4(11).

Kostek, M , Hubal MJ and Pescatello LS. 2011. The Role of Genetic Variation in Muscle Strength. American Journal of Lifestyle Medicine 5: 156-170.

Lango Allen H and numerous others. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature, 467:832-838.

Luciano M, Miyajima F, Lind PA, Bates TC, Horan M, Harris SE, Wright MJ, Ollier WE, Hayward C, Pendleton N, Gow AJ, Visscher PM, Starr JM, Deary IJ, Martin NG, Payton A.

2009. Variation in the dysbindin gene and normal cognitive function in three independent population samples. Genes Brain Behav. 8(2):218-27.

Mitaki S, Isomura M, Maniwa K, Yamasaki M, Nagai A, Nabika T, Yamaguchi S. 2012. Impact of five SNPs in dopamine-related genes on executive function. Acta Neurol Scand. doi: 10.1111/j.1600-0404.2012.01673.x. [Epub ahead of print]

Moran CN, Vassilopoulos C, Tsiokanos A, Jamurtas AZ, Mark ES Bailey, Hugh E Montgomery, Richard H Wilson and Yannis P Pitsiladis. 2006. The associations of ACE polymorphisms with physical, physiological and skill parameters in adolescents. European Journal of Human Genetics 14: 332–339.

Plomin R, Turic DM, Hill L, Turic DE, Stephens M, Williams J, Owen MJ, O'Donovan MC. 2004. A functional polymorphism in the succinate-semialdehyde dehydrogenase (aldehyde dehydrogenase 5 family, member A1) gene is associated with cognitive ability. Mol Psychiatry. 9(6):582-6.

Rizzi TS, Arias-Vasquez A, Rommelse N, Kuntsi J, Anney R, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Ruano D, Van der Sluis S, Markunas CA, Garrett ME, Ashley-Koch AE, Kollins SH, Anastopoulos AD, Hansell NK, Wright MJ, Montgomery GW, Martin NG, Harris SE, Davies G, Tenesa A, Porteous DJ, Starr JM, Deary IJ, St Pourcain B, Davey Smith G, Timpson NJ, Evans DM, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Faraone SV, Franke B, Posthuma D. 2011. The ATXN1 and TRIM31 genes are related to intelligence in an ADHD background: evidence from a large collaborative study totaling 4,963 subjects. Am J Med Genet B Neuropsychiatr Genet. 156(2):145-57

Singh KK, De Rienzo G, Drane L, Mao Y, Flood Z, Madison J, Ferreira M, Bergen S, King C, Sklar P, Sive H, Tsai LH. 2011. Common DISC1 polymorphisms disrupt Wnt/GSK3β signaling and brain development. Neuron. 72(4):545-58.

Smith CT, Boettiger CA. 2012. Age modulates the effect of COMT genotype on delay discounting behavior. Psychopharmacology (Berl). [Epub ahead of print]

Söderqvist S, McNab F, Peyrard-Janvid M, Matsson H, Humphreys K, Kere J, Klingberg T. 2010. The SNAP25 gene is linked to working memory capacity and maturation of the posterior cingulate cortex during childhood. Biol Psychiatry. 68(12):1120-5.

Williams AG, Rayson MP, Jubb M, World M, Woods DR, Hayward M, Martin J, Humphries SE, Montgomery HE. 2000. The ACE gene and muscle performance. Nature. 403(6770):614.

Zhang JP, Burdick KE, Lencz T, Malhotra AK. 2010. Meta-analysis of genetic variation in DTNBP1 and general cognitive ability. Biol Psychiatry. 68(12):1126-33.