

## REVISED VERSION:

### Providing education to adolescent girls in China - The power of water

Yasheng Maimaiti and W S Siebert, Birmingham University

August, 2008

#### Abstract:

This study investigates girls' secondary education dropout rates, and brings forward a novel variable : access to water. We hypothesise that girls' education suffers when their greater water need for female hygiene purposes with the onset of periods is not met. For testing we use six waves of the China Health and Nutrition Survey, 1989 - 2004. We find that the onset of periods is indeed associated with a marked increase in the secondary school dropout rate, but only for girls with poor access to water.

#### 1. Introduction

“If I had to pick the one thing we must do above all else to improve the world, I would say: ‘educate girls’ ”, Shashi Tharoor, India's contender for UN Secretary-General (2007, 165).<sup>1</sup>

In the developing world, the education of girls falls far behind that of boys, with widespread evil ramifications, reducing child health and womens' earnings (see, e.g., Schultz, 2002), and even women's role within the political system, since there is a strong link between education and political participation (Hannum and Buchmann, 2005). It is therefore important to find out why women fail to be educated properly in developing counties, and our paper seeks to make a contribution with a special dataset from rural China.

The paper starts from the observation that girls fall behind when the household has poor access to water. Figure 1 demonstrates this fact, using a cross-country dataset. Building on this link, we hypothesise that girls' education suffers when there is a lack of adequate water and sanitation both at home and at school. Lack of clean water for girls to wash themselves during their periods can lead to them staying away from school (Burrows et al, 2004). It is also possible that lack of water affects girls more than boys because girls “traditionally” do the water-carrying, which takes time out of school. Thus, we test whether, and to what extent, lack of water affects the education of girls more than boys. We find large effects.

The argument that lack of water has particular adverse effects on girls' education has been put forward before (Bista, 2004, Nahar, 2006; Kirk and Sommer, 2006; Singh, 1999; Snel, 2005, Burrows et al, 2004), but data limitations have prevented a test of the magnitude of the effect. We use six waves of the China Health and Nutrition Survey (CHNS), 1989 – 2004, to make the test. The CHNS dataset is well-adapted for this inquiry, because it has information on the household's access to water, as well as detailed information about time spent on household tasks such as water-carrying, and information about the onset of periods. It also has variables measuring family resources and income, which are important determinants of education, and clearly need to be controlled for the test. Admittedly, the dataset does

---

<sup>1</sup> We are grateful to Charles Diamond, MD of Econostat, for suggesting this quotation.

not contain information about availability of clean water at the school which respondents attend, and school water facilities can counteract bad home facilities (Nahar, 2006). Nevertheless, any finding of impacts from household access to water is of interest in itself.

We restrict the analysis to children in rural areas, since access to water is primarily a rural problem in China: according to the CHNS data, over 90% of families in urban China have direct access to clean (safe) water compared to only 48% in rural areas. However, as of 2005, nearly 60% of the Chinese population live in rural areas. Clearly, our investigation will be relevant to women's education in other rural communities in Africa, Asia, and South America to which most of the literature on water and sanitation currently relates.

As we will see, different estimation methods yield virtually the same results which support our hypothesis. We find that poor household access to water has a significant adverse impact on girls' education, and the impact only arises after the onset of their periods. As for the boys, poor access to water has no impact after controlling for income.

The paper is organised as follows. A literature review is presented in section 2. Section 3 describes relevant costs associated with poor access to water. Section 4 presents the model and its empirical specification. We describe the data in section 5. The empirical results are given in section 6. Concluding remarks appear in section 7.

## **2. Gender education gaps in China**

In China, primary and secondary education takes 12 years to complete, divided into primary, junior secondary and senior secondary stages. In general, primary education lasts 6 years. At junior secondary stage, most have 3 years schooling. The 9-year schooling period in primary and junior secondary schools pertains to 'compulsory' education. General senior secondary education lasts a further 3 years (Yang, 2006).

In secondary schools in rural China, substantial gender gaps exist which has generated much research (Connelly & Zheng 2003, Song et al. 2005, Brown & Park 2002, Hannum 1998, 2005). The literature on the reasons behind lower female educational attainment can be categorised as follows:

Opportunity cost: Some researchers argue that girls drop out of school, because their opportunity cost of staying in school is higher than boys. According to Li and Tsang (2002), in the past two decades the transition to a market oriented economy has allowed many privately owned enterprises to hire a large number of young female workers with limited education in the manufacturing and service sectors, especially in the booming coastal cities. Furthermore, rural villages and towns have developed various small-scale factories and enterprises that hire young women with limited education. These developments raise the opportunity cost of sending girls to school considering the earnings that they must forego. Connelly & Zheng (2003) and Song et al. (2005) also argue that girls have a higher opportunity costs due to the rise of light manufacturing jobs. In addition, Knight & Li (1996) argue that girls' education has a higher opportunity cost, since "traditionally" girls are family helpers. Daughters can

help parents with household chores and thus parents are able to work and generate more pecuniary income.

We have detailed information on children's time spent on household work and market work. We will test if working time has different impacts in the education models for girls and boys, and we will also test if the elasticity varies among girls who have good and bad access to water.

Household income and spending on education: this variable is important for education, as theorised by Gary Becker in his work on the economics of the family (Becker, 1981). However, the existing literature didn't come to a solid conclusion about the effects of income and spending on the creation of gender gaps in education in rural China. Brown & Park (2002) show the importance of household income in determining educational outcomes, but there is no indication in their empirical findings that poverty affects girls more severely than boys. Connelly and Zheng (2003) find county per capita income has larger effects for girls than boys for initial enrolment of primary schools (the significance of the difference is not provided), but the difference of the effect has disappeared in junior secondary school where the real gender gaps of school enrolment start to appear. In addition, Song et al. (2005) didn't find any statistically different coefficients in terms of household educational spending on girls vs. boys. Yueh (2006) using data from China's urban household survey, finds that household income seems not to matter for primary school enrolment, but there are conflicting effects for boys and girls at the various secondary stages. So the final outcome is unclear. Nevertheless, household income is important for our investigation to ensure that our access to water variable is not simply picking up the poverty of families which have poor access to water access, and therefore we lay emphasis on controlling carefully for it.

Future earnings and dependence: Song et. al. (2005) argue that patrilocal marriage traditions in China mean that the long term returns on investments in daughters are more likely to be realised by marital, rather than natal, families, while the reverse is true for sons. Thus, the education of a son is likely to be perceived as a necessary investment for support in old age. Admittedly they find that educating males appears only to bring small long term benefits in terms of household income - an extra year of education raises future expected income by only 0.9%. Still, these benefits are positive, and should predispose parents to prefer to educate sons. Li & Tsang (2002) also argue that, in rural Chinese society, married daughters are expected to contribute to the husband's family - which again means a son's education should be more profitable.

These arguments might imply lower education chances for girls from poor families which are more sensitive to possible losses due to daughters moving away. Certainly we will contrast rich and poor household below. Moreover, the arguments will drop away if our new variable, access to water is does a good enough job of explaining gender gaps.

Family characteristics: higher status of parental education and occupation is in general expected to have a positive impact on children's education, because educated parents in good jobs have direct experience of the benefits of schooling, which may make them develop a 'taste' for their children's schooling (Song et al., 2005, Emerson & Souza, 2007). Of course, higher parental education and better occupation will also

mean higher household income, so our control for these variables can also serve to control for income.

Siblings: Different numbers of siblings and sibling structure may also have different impacts on parents' education decisions (Connelly & Zheng 2003, Yang 2006, Tsui et al. 2002). While the findings remain mixed, we agree that detailed controls for sibling structure are necessary. In particular, if older girls "traditionally" help at home, we might expect to see an elder sister to be advantageous for a girl's education, but not a boy's.

Geographical location: Children living in remote areas lack nearby schools, adequate transportation and information. All these may have negative impacts on their school enrolment. And the negative impact may be bigger for girls, especially for their secondary school enrolment. However, Connelly & Zheng (2003) find no evidence that living in a hilly county has a significant negative impact on school attendance of girls in all school levels. Li and Tsang (2002) assert that safety concerns clarify why parents from remote rural settings have lower educational expectation for their daughters – schools are simply too far away, given the poor transport, giving boys the advantage. But they provide no direct measure to test this hypothesis. We discuss below whether our access to water variable is picking up this location effect – since poor access is likely to overlap with remote geography. However, our controls for family income, and also for travel time to school will hopefully sweep out this effect.

'Culture': Many researchers address traditional and cultural aspects in China that have potential impacts on the school enrolment of boys and girls. Li & Tsang (2002) describes how "families without sons are recorded as having died out". This rigid lineage system, along with the economic, social and political advantages of educating a son discussed earlier, generates the concept of pro-son bias in schooling decisions. Song et al. (2006) argue that a son's education is more of an 'investment' good in rural China, whereas a daughter's is often taken as 'consumption', even 'luxury' good. Moreover, Knight and Li (1996) find that while traditional values favouring boys' education appear to have been eroded in urban areas, they have not in rural areas – which are our focus.

The difficulty with arguments relying on culture is that we need to know where the "culture" itself comes from. As Yueh (2006) argues, cultural traditions are often the historical product of practical necessity, so that rational acts under one set of circumstances, such as in rural China, will change when the context is altered, as in urban China with a different set of household needs and constraints. In any case, the pro-son bias, if it exists in the rural areas, will exist irrespective of access to water. Any finding that girls with good access to water are not disadvantaged compared to boys will cause arguments based on culture to fail.

In sum, previous research aiming at addressing gender gaps in education in China has come to mixed conclusions. The CHNS provides a large dataset to re-investigate this inquiry, and with detailed information on individual, household and community characteristics relevant to our research. In particular, with the CHNS, we have special data on access to water and personal hygiene including the onset of periods. Both factors are likely to be important for girls' schooling, as we now describe.

### **3. New variables: access to water and girls' periods**

Much economic and educational research has been conducted in terms of access to clean water and its direct and indirect impacts on girls' schooling in other developing countries where having access to clean water remain a big social problem (Behrman et al. 1999; Holmes 2003; Colclough et al. 2000; Hill & King 1995; Davison et al. 1992). In rural China, access to clean water has also been regarded as a major social issue. For example, at least 300 million rural residents in China had no access to safe, clean drinking water in 2006 (source: Xinhua, Wednesday, August 30, 2006). At the same time, the link between water and girls' education has not been made clear. Here we categorise the literature which links access to water with girls' education.

First, time costs generated by fetching water may have adverse impacts on girls' schooling as girls are considered to be the main water-fetchers in many African and South Asian Countries (Hill & King 1995; Behrman 1997). The production of schooling involves time costs, so does fetching water. But are girls in rural China also considered to be the main water-fetchers? If that is the case, is the distance to fetch water long enough to affect their time for schooling?

Secondly, health costs generated by poor access to safe and clean water are expected to be more for girls. There are many studies that report significant associations between child health and child schooling performance (see for example, Behrman 1996). We assume households allocate their resources rationally (Foster 2002). Assuming also that marginal benefits of educating a son and a daughter is the same, if the marginal cost (here including the health costs) of educating the daughter exceeds that of boys, parents will invest more in their sons' schooling. Girls' health will suffer more if proper personal hygiene is not in place during their periods (see Poureslami & Osati-Ashtiani (2002), Beek et al. (1996), Dagwood (1995) and Secerino & Moline (1995)).

Finally, psychic costs may arise for girls during their period if they are unable to clean themselves (Bista 2004, Nahar, 2006; Kirk and Sommer, 2006; Singh, 1999; Snel, 2005). In schools, toilets may be totally absent or few in number, with broken doors or defective water supply and sewerage (El-Gilany 2005, Behrman et al. 1997). Rose Lidonde (2005) has explained the problems that girls face in detail in an African context. Snel & Shordt (2005) also claim that school drop-out rates and low literacy levels, especially among adolescent girls, can be attributed in part to inadequate sanitation conditions in schools. In addition, Cairncross et al. (1998) find that a school sanitation programme in Bangladesh increased girls' enrolment by 11 per cent.

The arguments above boil down to the belief that that girls' education suffers when greater time, health and psychic costs associated with poor access to clean and safe water are generated. Burrows et al (2006, 14) put it well:

“Lack of adequate water and sanitation both at home and school prevents girls from attending school when menstruating. Girls have a sense of being unclean when there is little clean water to wash themselves, and this can lead them to stay away from school. Also there are rarely private facilities at school where girls can go to the toilet or wash the rags they use during their periods. They can also pick up infections if the water they use to wash rags is dirty, leading to more time off school”

Figure 1 provides a simple model of our hypothesis. We picture the cost curve as taking a step upwards with the onset of periods in families with poor access to water. For given education demand (D), girls in these families will choose only A years of schooling. Girls in otherwise similar families, but with good access to water – and boys – will choose B years of schooling. Of course, family income will shift both demand and supply curves, and we control for income. Ability is more of a problem, but should not be correlated with period or access to water, our variables of interest.

Table 1 provides motivation for the hypothesis. We take the enrolment rate for the rural population in age groups 6 to 18, and distinguish four groups of females by whether or not access to water is good (defined in more detail below), and whether periods have begun. For comparison, we also show results for boys, distinguishing those with and without good access to clean and safe water.

As can be seen, the initial enrolment rate is high for all groups, but girls with poor access to water (no access to tap water) tend to drop out, particularly after periods have begun. Thus, the school enrolment rate of girls aged 11-15 whose periods have begun, and who have poor access to water, is only 60%, whereas girls without periods, and boys, have an enrolment rate of 86% or more, irrespective of access to water. The Kaplan-Meier survival curves in Graph 1 also shows that girls with poor access to water have a much lower survival curve than the other three groups, having only about a 20% chance of surviving until period 9, which is the end of junior secondary school. These basic data seem to back up our hypothesis, and we will now test it using appropriate statistical methods.

#### 4. The statistical model

For initial analyses below we use survival models (Hosmer & Lemeshow, 1999, Kalbfleisch, & Prentice 1980), which estimate the probability of dropping out of school conditional on school enrolment until the previous grade. A disadvantage of this approach is that the the dependent variable which is schooling duration is highly correlated with the pupil's age, and also the date of onset of periods, so it is difficult to exert a fine control of age when entering the period variable. Subsequently, we therefore present results using a logit model, based on whether or not the individual is currently at school. This model allows better control of age and periods, and can be estimated using the Generalised Estimating Equations (GEEs) technique (Hardin and Hilbe, 2003) which is more robust. A further model uses as dependent variable the individual's accumulated schooling relative to average schooling for individual's of that age, in that county (age-adjusted schooling). This specification allows perfect control for age, in that we consider the effects of access to water within exact age groups.

Looking first at the survival models, the simplest approach is the Cox proportional hazard model, in which the hazard rate,  $h_i(t)$ , that is the exit probability from schooling, conditional on being at school up to time  $t$ , is specified as follows:

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \dots + \beta_k x_{ki}) \quad (1)$$

where  $h_0(t)$  is the baseline hazard, and the  $x$ 's are covariates. Here the baseline hazard is not given a parametric form (Cox & Oaks, 1984). However, we modified this approach from the outset, since tests rejected Cox's proportional hazards assumption (for girls: df=24; Chi-square=44.11; p=.01. for boys: df=24; Chi-square=44.48; p=.01).

Consequently we use Weibull accelerated failure time (AFT) models which give a parametric specification to the baseline hazard.<sup>2</sup> The AFT coefficient gives the relative probability of duration of schooling, conditional on being at school up to time  $t$ , and is specified as follows (Cameron & Trivedi, 2005, 592):

$$\lambda(t | \mathbf{x}) = \lambda_0(t \exp(-\mathbf{x}'\boldsymbol{\beta})) \exp(\mathbf{x}'\boldsymbol{\beta}) \alpha \quad (2)$$

where  $\lambda(t | \mathbf{x})$  is the hazard rate,  $t$  is survival time, the  $\mathbf{x}$ 's are covariates, the  $\boldsymbol{\beta}$ 's are time ratios which need to be estimated, and  $\alpha$  is a correction for heterogeneity (Hosmer and Lemeshow, 1999, 318), distributed as gamma with mean 1 and variance  $\theta$ . Here, the time ratios can be interpreted as survival time multipliers. For example, a time ratio of 1.5 means that, conditional on being at school up until the instant past moment, the relative time of surviving the school is 1.5 times (50%) longer if the independent variable increases by one unit. Time ratios greater than one correspond to positive coefficients and time ratios less than one correspond to negative coefficients.

Looking next at the logit type of model, while this model has the advantage of allowing for better control for age, there is the disadvantage that our panel data are quite unbalanced (see below). Consequently, using a fixed effects approach would result in losing many observations – this problem also makes random effects and fixed effects specifications difficult to compare (since they are based on such different observation numbers). A solution is the GEEs approach, first introduced by Liang & Zeger (1986), which gives unbiased estimation of parameters and standard errors, even when the correlation structure is misspecified (Ghisletta and Spini, 2004). We use this approach both when analysing the probability of school enrolment (logit), and the age-adjusted schooling variables.

Our independent variables of interest are access to water, and the onset of the period. As regards the access to water variable, this effect is obviously allowed to be different for boys and girls. For both sexes, it is natural to expect poor access to water to decrease the likelihood of children's school enrolment, not only because of the general water related health problems, but also because of the increased time required to fetch water. In this respect, poor access to water equally reduces both boys' and girls' schooling. However, we expect poor access to water to have a worse impact on girls' schooling after periods have begun, due to the hygiene related economic and psychological problems that girls face as described above. These considerations point to an interaction between the access to water and period variables.

We also control for household income, which is an important variable, as explained above. In our case, controlling for household income is even more important, given the fact that high income is likely to be associated with good access to water. Hence, we need to measure income well in order to separate its effect from the effects of access to water. While measuring household income is difficult in a rural society where much income does not arise through the market, the CHNS offers good data. In particular, (see below) the CHNS asks respondents about a comprehensive set of income generating activities. To supplement this measure of household income, we also have variables for father's and mother's education and occupation. Generally,

---

<sup>2</sup> The Weibull has the smallest Akaike Information Criterion values (1460.6 is the result for girls) compared to Log-logistic (1479.4) and Log-normal (1518.9).

higher status of parents' education and occupation always go together with the wealth of the family.

A further consideration is children's work, both market and household, which needs to be included to control for the alternative uses of children's time. A possible problem is that children's market and household work will be chosen by the household jointly with schooling, so these variables are endogenous. We need to find instruments for these variables which of course is difficult. Fortunately, however, even in rural China few children report that they help much with market work or household work (see below for details). Hence, when we experiment by dropping the two work variables from the regression, similar coefficients remain for most of the other explanatory variables, with the impact of the work variables mainly changing the impact of the age dummies.

## 5. Data

The data used in this study come from the Chinese Health and Nutrition Survey (CHNS), jointly conducted by the University of North Carolina and the Chinese Academy of Preventive Medicine, Beijing. The CHNS is designed to examine "how the social and economic transformation of Chinese society and family planning programs implemented by national and local government affect the economic, health and nutritional status of its population" (CHNS, 2007). The survey drew a sample from nine provinces: Heilongjiang, Liaoning, Henan, Shandong, Jiangsu, Hubei, Hunan, Guizhou, and Guangxi. These provinces stretch from the North-East to the South-West, and vary substantially in geography and economic development (Yang, 2006). Households in each province are selected using a stratified multistage cluster design that includes approximately 20 households in each of some 190 urban and rural communities.

The CHNS has interviewed households six times between 1989 and 2004. While this period is quite long, the main change in education law, requiring 9 years of compulsory education, was in 1986 (Yang, 2006), well before our first survey. Hence, we will simply rely on intercept dummies for each survey to pick up trend effects.

Our analysis uses information only from rural settings, where "urban" and "rural" are defined according to household registration. For each wave of the survey, we identify children graded 1-12 and usually aged 6-19. We do not consider tertiary education, after age 19, since different factors such as marriage enter the schooling decision. In addition, our hypothesis relates to the interaction of poor access to water and the onset of periods for girls. In our sample, the mean age of onset of periods is 12.7 (compare 13.5 in Singh (1999), and 12.9 in El Gilany (2005)), and thus, with the 6-19 age span, we have adequate observations from both before and after.

In table 1, the means and standard deviations are shown for variables used in the analysis. About 72% of the individuals in the sample are at school at the time of survey, 47% of the samples are girls. And 49% of the girls for whom we have the period information have begun periods.

The access to water variable is derived from the question 'how does your household get drinking water?'. In the original sample, 30% of the children had tap water at home, 17% tap water in the courtyard, 34% well water in the courtyard and

19% other water sources outside their courtyard. Thus, 53% of the children have no access to tap water, which we use as our definition of “poor” access to water.

The CHNS provides a detailed per capita income estimate for rural households, which is not usually available from other sources. Gross household income in cash or kind is created for different categories and then expenses are deducted to create a net income value, deflated using the appropriate price deflators. The main problem is that much income is not received via the market, and a market value has to be imputed. To measure income in-kind, the CHNS relies on the respondent’s (usually the household head’s) estimation about the market value of the goods consumed and received as gifts. For home gardening income, the total value of household food consumed at home or sold is measured. Income from farming, raising livestock/poultry; collective and household fishing; and the value of income from other household business is obtained by same calculations. The CHNS also takes into account welfare subsidies including housing subsidy, child care subsidy and gifts. In our sample, the mean per capita rural household income is 1,225 *yuan* (1988 community CPI) for the period of 1989-2004. This figure is generally in line with the estimates from other sources. For example, mean per capita household income for rural residents in China is 1,067 *yuan* (1990 CPI) for 1987-2001 in Benjamin et al (2005).

The dataset also allows us to use most of the conventional control variables for determinants of children’s schooling. Educational qualifications of parents are grouped into four categories: 0 being no qualification (18% for fathers and 43% for mothers), 1 being graduation from primary school (28% for fathers and 23% for mothers), 2 from junior secondary school (37% for fathers and 24% for mothers) and 3 from senior secondary school or college (17% for fathers and 10% for mothers). We also include parents’ occupational status, grouped into 4 categories: 1 being high occupational status (6% for fathers and 1% for mothers); 2 medium occupational status (16% for fathers and 10% for mothers); 3 to no wage work (mostly farmers, and 64% for fathers and 71% for mothers) and 4 to others (including unemployed, and 11% for fathers and 18% for mothers).

As for sibling structures, we see that 31% households have a single child, 41% have two children and 28% have more than two children. We also account for sibling structure in the households who have two children, to measure the overall impacts of ‘gender bias’ or ‘son preference’ if any. We find that within this group 12% of children have one older brother, 9% one older sister, 10% one younger brother; and 10% one younger sister.

Finally, we also have household and market work variables measured as hours worked per day to capture the impact of work opportunities on children’s schooling. But, in fact, only 17% of girls aged 13 or over report more than 1 hour/day household work (including washing clothes, cooking, laundry and caring for younger brothers and sisters), and much fewer boys, about 4%. In addition, again 17% of girls aged 13 or over report more than 1 hour/day market work (farming, livestock raising, working in local business and gardening), and 10% of boys. Admittedly, these figures show girls are still doing a relatively greater amount of household work than boys, but their time doing market work is similar. Below, we control for these variables, but expect their overall impacts to be small for education choices.

## 6. Empirical results and discussion

Our empirical strategy is as follows. First we test if poor access to water has any adverse impact on children's schooling. We use Weibull AFT models for such test. Secondly, we investigate the reasons as to why poor access to water matters for girls' education. Here we will use GEEs for logit analysis of school drop out, and age-adjusted school attainment. In the final step, we will use community level panel data to quantify the impact of poor access to water on school enrolment of children.

### 6.1. Does poor access to water matter?

Table 3 gives the results of Weibull AFT model controlling for individual specific heterogeneity. Here we use first three water categories (tap water at home, tap water at courtyard and well water at courtyard) as good access to water and other water sources outside the courtyard as poor access to water. There is a striking difference of the impact of poor access to water on girls' and boys' schooling. As can be seen, after controlling for family income, parental education and occupational statuses, time of household and market work, sibling structure and other conventional control variables, we find that the impact of poor access to water on boys' schooling has disappeared, while the impact is strong and consistent in the girls' model. We also note that the girls and boys regressions can be pooled:  $\chi^2(36) = 45.26$ ;  $\text{Prob} > \chi^2 = 0.14$ . And in the pooled regression, the cross product of poor access to water and female remains big and largely significant.

Table 4 checks whether we can use a simplified clustered income variable, which will simplify later analysis greatly. Here, the high and low income categories are based on k-means clustering using per capita household income, and the parental job status and educational qualifications listed in Table 3. The cut-off point of per capita household income is set at its median value, while job status and educational qualifications variables are kept with their four original categories. This process classifies the households 9572 "poor" and 2962 "rich". As is shown in the table, the impacts of poor access to water remain virtually the same. The results confirm that we can use the clustered income variable to construct cross products to test whether the impact of water and periods changes by income.

A possible alternative reason for the adverse effects of poor access to water for girls is that poor access proxies for extreme poverty. However, for this argument to hold, poor access to water would have the same negative impact on girls prior to the onset of periods, as afterwards, or on boys. Another argument is that poor access to water may relate to remote villages, and the water variable simply picks up the extra cost of going to school. But again, if this were the case, girls prior to their periods, and boys, would be similarly affected. Nevertheless, we will use appropriate methods and variables to test if these arguments hold.

The CHNS dataset introduces the time to school variable from the wave 1997. We include the variable in the schooling duration models using 3 waves of the survey data (waves 1997, 2000, 2004), to test whether distance to school adversely affects school attendance. In fact, the effects are small (results not shown), whereas the impact of poor access to water remain large. Hence, we conclude that the impact of water does not pick up the impact of backward geographical location. Similar to the findings of

Connelly & Zheng (2003), we find no significant location effect on children's schooling.

We also test if the impact of water differs by family income and the results are shown in table 5. We divide the girls into poor and rich families by clustered income, and run separate regressions for these two groups. Interestingly, we find the adverse impact of poor access to water remains large irrespective of family income. In fact, we find that the impact of poor access to water is not statistically different in the two regressions:  $\chi^2(1) = 0.44$ ;  $\text{Prob} > \chi^2 = 0.51$ . The implication here is that, since the adverse impact of poor access to water holds in the "rich" group, the access to water variable cannot simply be picking the effect of poverty. Thus, poor access to water has an impact, irrespective of location and poverty, and the impact is only on girls.

## 6.2. Why does poor access to water matter for girls' schooling?

We now turn to our main hypothesis, the role of periods in explaining the adverse impact of poor access to water for girls. Other possible reasons for the impact of poor access to water, that it proxies for longer distances to school, or extreme poverty have been discussed, and appear to be ruled out. There remains the possibility that for "cultural" reasons it is mainly the girls' chore to fetch water, and this factor makes poor access to water adverse for girls. We will first examine whether the impact of poor access to water varies before and after the onset of periods to test our main hypothesis. We will use our age-adjusted schooling models to test the cultural girls' chore explanation.

In the following regressions we will use 4 water categories, namely, tap water at home (water 1); tap water at the courtyard (water 2); well water at courtyard (water 3) and other water sources outside the courtyard (water 4). Using these graduated water categories permits more detailed water-period cross-products. As noted above, we now use school enrolment (1 is enrolled at school at the time of survey, 0 not) as our dependant variable, since this variable is less collinear with the period variable we are going to exploit. Nevertheless, the overall concept remains unchanged – the girls who drop school early accumulate less years of schooling.

Table 6 shows striking evidence that poor access to water has a more adverse impact after girls reach puberty. In two different specifications (model 1 uses all 17 income-related variables, and model 2 uses the clustered income variable), the impact of poor access to water is significant and large for girls after the onset of their periods. The 'pure' impact of poor access to water disappears after we introduce the period variable and the period-water cross products. For example, girls who have begun periods and who do not have access to water in their courtyard (water 4) are 13% more likely to drop the school than those who have access to tap water at home. Those who have well water in the courtyard rather than tap water at home also do worse (-9%).

The cultural chore argument competes with our hypothesis by asserting that the results can be interpreted merely as 'age effects', since it is the older girls who reach puberty and are then fetch the water. Hence, when the access to water is poor, the older girls suffer more than the younger ones simply because they fetch more water. In order to test the validity of this argument, we construct a new dependent variable – age-adjusted schooling. The adjustment proceeds by subtracting from each respondent's schooling the average years of schooling for girls of her age and in her county. This

variable permits us to test whether the impact on schooling of poor access to water varies for girls with and without periods, holding age the same.

Table 7 gives the results, and reconfirms our hypothesis that the impact of poor access to water comes through the onset of period for reasons other than the time spent fetching water. In other words, the results show that poor access to water decreases school attainment of girls who have started periods compared to those who have not, of exactly the same age – and thus with the same putative water-carrying responsibilities.. Having no tap water (water category 3 and 4) is again proved to be detrimental even for girls ‘age-adjusted schooling’.

We also introduce further cross products in Table 7 to test whether the impact of income varies by good/poor access to water and before/after the onset of periods. The 3 factor (income-water-period) cross products test joint marginal impacts. In the first specification, we find that relative income has a strong positive impact (0.48) on girls’ age-adjusted schooling. However, in the second specification we find that most of the relative income effect is via the period variable (0.67), implying that high incomes help girls primarily after the onset of periods. But the third specification gives the clearest picture: we see that relative income helps most (0.82) when water access is poor and periods have begun. Note that in all the different specifications, we see consistent adverse impacts of the water-period interactions.

The other variables in the regression models listed in table 4 – 8 behave much as expected, and generally are not significantly different as between boys and girls, and so cannot explain the gender education gap. Several of the other proxies for household income such as parental educational qualification and occupational status seem to have different impact on boys’ schooling. However, likelihood ratio tests cannot reject the hypothesis that the income coefficients are statistically different between boys and girls. In general, then, increases in household resources significantly increase schooling duration for both boys and girls. As for other variables, we see that sibling structure variables have neutral impacts on children’s schooling in most of the specifications the only exception is being that having an older brother always hurts girls’ and boys’ schooling (in line with the findings of Yang (2006)), but the impact is not statistically different between genders. This result validates our suspicion that after controlling for socio-economic factors, so called ‘pro-son’ bias does not exist, or at least its existence is insignificant. Furthermore, household work and market work opportunities have significant, though small, negative impacts on girls’ and boys’ schooling. At the same time, we find that the impacts for the different genders are the same statistically.

In short, although the impacts of household income, parental job status and parents’ educational qualifications have strong impacts on children’s schooling, their impacts are not statistically different between genders, which means that they are not the main factors that generate the gender gaps in education. And household and market work opportunities do not seem to have gender specific impacts on children’s schooling either.

### 6.3. How great is the impact of poor access to water on girls' schooling?

So far we have confirmed that poor access to water affects girls schooling, and the impact comes through the interaction with periods. Now we are interested in quantifying the impact. Table 8 presents percentiles of some survivorship distributions for different scenarios for access to water and family income. Scenarios (1) and (2) show the impact of access to water holding other control variables at their mean, and we see that with poor access, girls at the median lag on average two years (11.2 – 9.2), whereas the impact of poor access is only marginal for boys (10.6 – 9.9).. The two year average difference in school attainment can be interpreted as about 35% less schooling considering the average school attainment of girls in rural China is about 5.6 years across 1989 – 2004.

Scenarios (3) and (4) vary household (clustered) income, holding other control variables at their mean. Here we see that being in the poor family category pulls both girls and boys down about the same amount, demonstrating that income has the same effect on boys and girls. Taking the median, both boys and girls accumulate about two years (35%) less schooling if the family is poor than if it is rich. Income is important, but cannot account for the gender gap.

In rows (5) to (8), we vary both access to water and family income category holding other control variables at their mean. In scenarios (5) and (7), we assume water access is good and allow the impact of income to vary. We find that girls and boys accumulate virtually the same years of schoolings in high income category (12 years at the median both for boys and girls) and in low income category (at the median, 10.2 years for girls and 9.9 years for boys).. However, assuming water access is bad – scenarios (6) and (8) - girls from both high income and low income families accumulate about one year's (18%) less schooling compared to the boys in the same categories. This fact again confirms that poor access to water poses special impacts on girls' schooling irrespective of family income.

Next, assume that income is high (scenarios (5) and (6)) and let the impact of access to water vary. Here we find that there is about a one year difference in school attainment between rich girls who have good access to water and who have not. In line with our findings, for boys from high income families, the poor access to water does not have any impact. Finally, assume income is low (scenarios (7) and (8)) and allow the impact of water to change. Interestingly, girls from poor families still outperform boys (10.2 to 9.9 years at the median) as long as the access to water is good, but fall behind (8.3 to 9.4 years at the median) if the access to water is bad. And even between poor family girls with and without access to good water there is a two year (35%) difference in school duration (10.2 to 8.3 years at the median). But, for boys from poor families the impact of different water sources remains small (9.9 to 9.4 years at the median). It is also interesting to note that wherever there is poor access to water (scenarios (2), (6) and (8)), girls always lag behind boys in all the percentile points. In the other situations either water access is good, so girls are not disadvantaged, or the family is rich, and can overcome the access to water problem for its daughters. Of course, unfortunately, it is the scenarios with poor access to water which is most prevalent in the developing world. In short, the different scenarios indicate two facts: water elasticity is big and consistent for girls irrespective of family income but negligible for boys. Income elasticity of schooling is big but similar in size across

different genders and the elasticity gets marginally bigger for both genders when access to water is poor.

The findings confirm that increasing family income should improve girls and boys schooling sizably, but would not improve the gender gap. But improving access to water has a very specific benefit for girls' schooling while it has very little impact on boys'. Improving access to clean and safe water can dramatically increase average years of schooling of girls (from one to two years – 18-35%), and reduces the gap.

As a final step, we develop some policy implications using community level water variables which are directly linked to government policy. We first derive community level variables (about 130 communities) using individual level data. We drop communities which have less than 10 observations and construct community average school enrolment rates of girls and boys for each wave (see Table 2, bottom panel). We use these variables as dependent variables. We also construct a set of community level explanatory variables based on community averages (Table 2). For policy analysis, we track changes of the water access variable across the whole span of years and derive its contribution to the gender gap.

As is obvious from Table 9, the community rate of poor access to water again poses no obstacle (0.00) to the community rate of school enrolment of boys after controlling for community income levels. However, there is a significant impact of poor access on community rate of school enrolment of girls after the onset of periods (-0.34) who have reached their puberty. A one percentage point change in community average access to good water can increase girls' school enrolment by about 0.34 percentage points holding community rate of girls' periods at the mean. From this estimation results we can compute the overall contribution of improved access to water on school enrolment. In particular, girls who have started their periods in rural China have increased their school enrolment by 18 percentage points during 1991-2004, and access to good water increased by 19 percentage points during the same period. Hence, holding other things equal, the overall contribution of improved access to water on these girls' school enrolment is 33% ( $=0.34*19/18$ ), quite a lot.

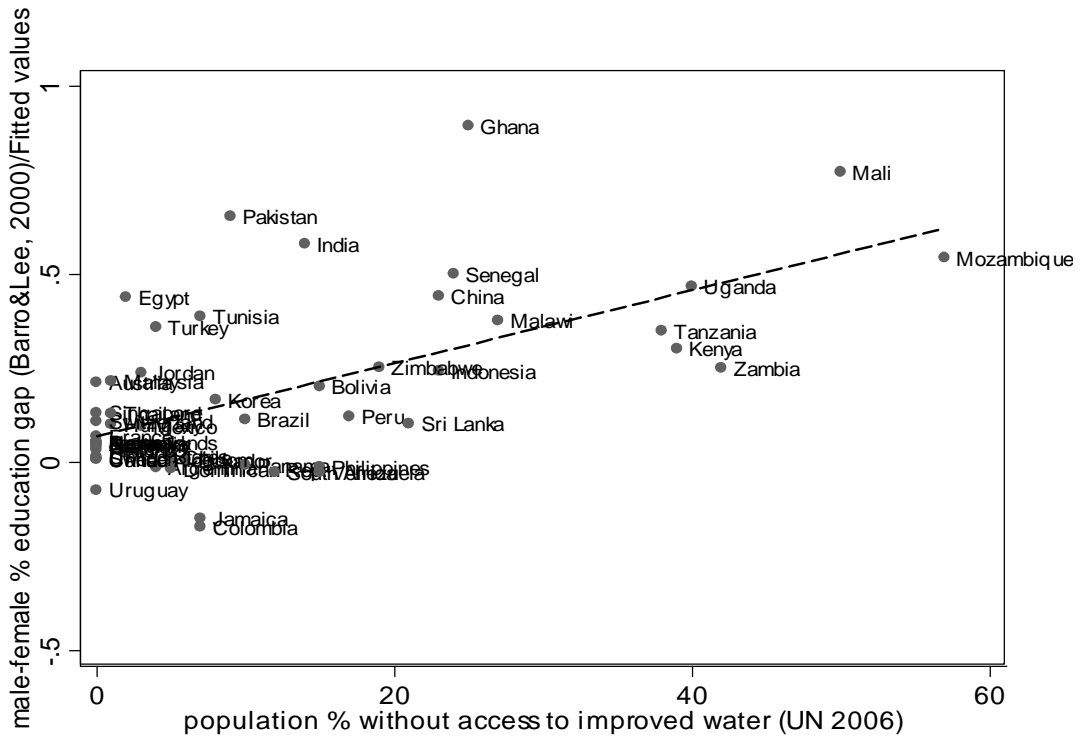
## **6. Conclusion**

Narrowing the gender gap in education is profoundly important for economic and social development. Women cannot play their proper role in public affairs unless they are educated as well as men. Democracy itself suffers. In rural China, there are gender gaps in school enrolment in secondary education. However, previous research addressing the reasons for the gap give mixed results. Our results are clear: poor access to water is what drives the education gap between girls and boys. Moreover, because we have data on the onset of periods, we can point to the reason why poor access to water has this adverse effect. It is primarily women who have periods that suffer the consequences of poor access to water. Our findings simply underline the fact that periods require water for cleanliness, without which school-going becomes very difficult.

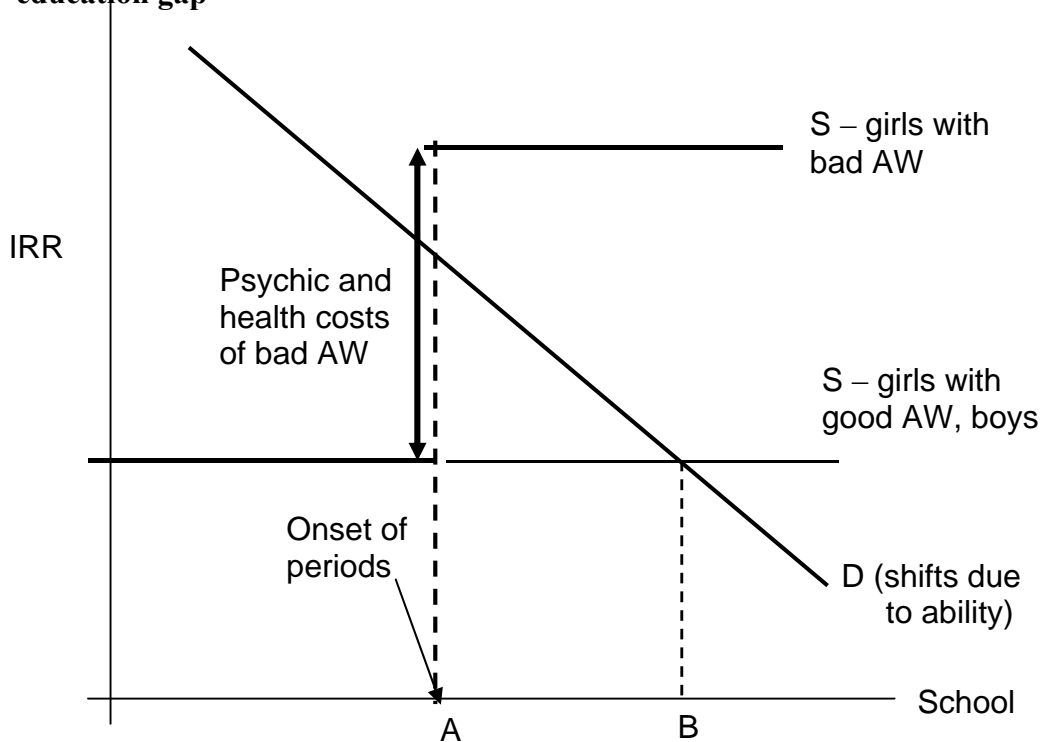
Our research gives estimates of the weight of the disadvantage that girls with poor access to water face. Not having good access to water has little effect on boys, or girls prior to their period. But, for girls after their periods, lack of water is a serious disadvantage, causing a 2 years fall in school duration (Table 8), or a 12% fall in the

probability of enrolment (Table 6). Our results have a striking policy implication. A major benefit of policies to improve water supplies may not be the obvious household or industrial benefit, but rather an unseen benefit, the improvement in the position of women.

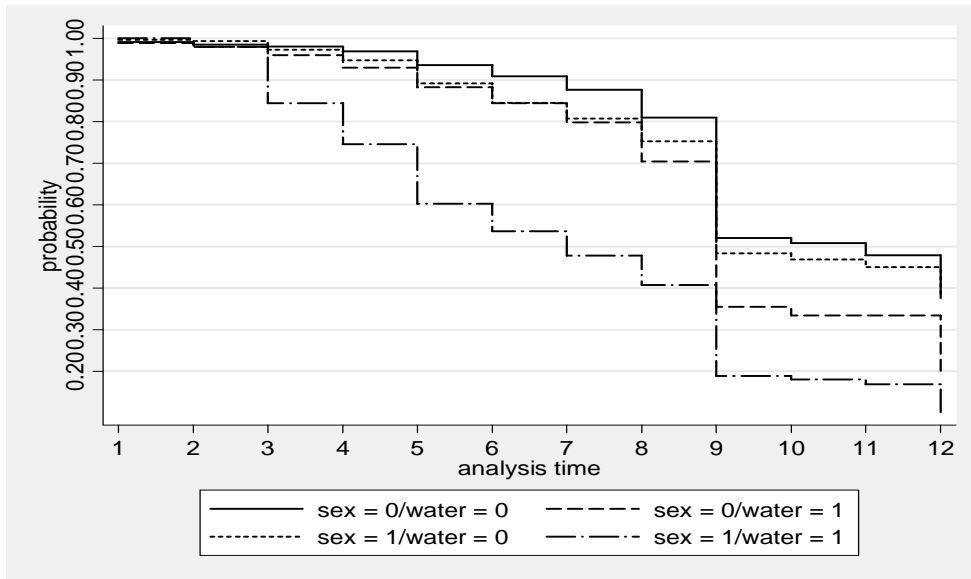
**Figure 1: Male-female education gap is higher in populations with lower access to water**



**Figure 2: Model of access to water × period interaction and male-female education gap**



**Figure 3: Survival distribution of schooling for children with good vs. poor access to water**



**Notes:** analysis time corresponds to accumulated years of schooling: 0-6 being primary, 7-9 junior secondary (the end of compulsory education), and 10-12 senior secondary school; boys = 0; girls = 1; good water=0; poor access to water=1 (for definitions, see Table 2); for girls, the analysis is restricted to those who have started periods.

**Table 1. Descriptive Statistics**

Variables	N	Mean	S D	Description
<b>Individual Level:</b>				
years of schooling	12871	5.64	3.00	accumulated years of schooling
school enrolment	13233	0.71	0.45	whether at school (=1) or not (=0)
gender	14094	0.47	0.5	gender (female=1)
period	5992	0.49	0.5	if having period (=1) or not (=0)
water 1	13860	0.30	0.46	having tap water at home
water 2	13860	0.17	0.37	having tap water in courtyard
water 3	13860	0.34	0.47	having well water in courtyard
water 4	13860	0.19	0.39	other water outside the courtyard
ln_income	14002	6.78	0.86	log of per capita net household income
mother's occupational status	14094	3.05	0.57	4 occ. status representing: 1= high; 2= medium; 3= farmer/no-wage work; 4= other/unemployed
father's occ. status	14094	2.86	0.72	same as above
mother's educational qualifications (EQ)	13085	1.00	1.02	4 educational qualifications representing: 0= no qualification; 1= primary school; 2= junior middle school; 3= high school/college
father's EQ	12840	1.54	0.97	same as above
clustered income	12534	0.24	0.42	clustered income 1=rich
single child	14094	0.31	0.46	single child at home
one sibling	14094	0.41	0.49	child having one sibling
two/more siblings	14094	0.28	0.44	having two or more sibling
household work –girls	6688	0.41	1.21	number of hours spent for household work per day last week
household work – boys	7406	0.09	0.45	same as above
market work – girls	6688	0.48	1.64	number of hours spent for market work per day last year
market work – boys	7406	0.42	1.51	same as above
age	14094	12.79	3.9	age of children
<b>Community Level:</b>				
girl_com	649	0.70	0.21	community enrolment rate of girls
boy_com	647	0.73	0.19	community enrolment rate of boys
water_com	647	0.17	0.30	community access rate to poor access to
income_com	649	0.27	0.31	community proportion of riches
work_girl_com	649	0.85	0.98	community work hours for girls / day
work_boy_com	649	0.51	0.66	community work hours for boys / day
child_com	649	1.95	0.48	comm. number of kids per household
age_com	649	12.8	1.36	comm. average age of children
period_com	649	0.50	0.22	community rate of menstruating girls among girls being surveyed

**Table 2. Average School Enrolment Rates % (Rural China, CHNS: 1989-2004)**

Age	Girls				Boys	
	Period	Good Water No period	Poor access to water Period	No period	Good Water	Poor access to water
7-10		.95 (.23)		.92 (.27)	.95 (.21)	.93 (.25)
11-15	.82 (.39)	.92 (.28)	<b>.60</b> (.49)	.86 (.35)	.91 (.29)	.86 (.35)
16-18	.35 (.48)		<b>.19</b> (.40)		.40 (.49)	.28 (.45)

**Notes:** see Table 2 for definitions of good/poor access to water. Standard deviations are reported in parentheses.

**Table 3. Weibull AFT Model with Heterogeneity**

Dependent variable: schooling duration

	Girls		Boys		All	
	Coef.	z	Coef.	z	Coef.	z
water	-0.17	-3.45	-0.02	-0.46	0.00	0.02
female					0.03	1.11
water × female					-0.18	-3.01
log income per capita	0.07	3.20	0.02	0.93	0.05	3.14
father - high occ. status (ref)						
father - medium occ. status	-0.14	-1.82	-0.06	-0.85	-0.10	-1.99
father - no wage work/farmer	-0.20	-2.73	-0.16	-2.72	-0.17	-3.77
father - other/unemployed	-0.29	-2.50	-0.17	-1.74	-0.23	-3.06
mother - high occ. status	0.46	1.27	0.16	1.01	0.23	1.50
mother - medium occ. status (ref)						
mother - no wage work/farmer	-0.16	-2.09	-0.05	-0.70	-0.11	-2.17
mother - other/unemployed	-0.13	-1.49	-0.02	-0.28	-0.07	-1.28
father - no school qualification	-0.25	-3.33	-0.17	-2.78	-0.22	-4.59
father - primary school	-0.19	-2.82	-0.18	-3.38	-0.20	-4.65
father - juniour middle school	-0.12	-1.92	-0.03	-0.64	-0.08	-2.01
father - high school/college (ref)						
mother - no school qualification	-0.24	-2.71	-0.30	-3.57	-0.27	-4.41
mother - primary school	-0.18	-2.06	-0.33	-3.94	-0.25	-4.18
mother - juniour middle school	-0.13	-1.47	-0.21	-2.64	-0.17	-2.83
mother - high school/college (ref)						
children's household work	-0.06	-4.19	-0.07	-2.08	-0.06	-4.96
children's market work	-0.10	-7.03	-0.09	-7.87	-0.10	-10.92
single child (ref)						
one older brother	-0.09	-1.57	-0.12	-2.31	-0.11	-2.84
one older sister	0.03	0.44	-0.08	-1.67	-0.04	-1.07
one younger brother	-0.06	-0.84	-0.19	-3.09	-0.13	-2.79
one younger sister	0.01	0.07	0.01	0.17	0.01	0.22
two/more siblings	-0.08	-1.55	-0.05	-0.98	-0.07	-1.99
age 6-11 (ref)						
age 12-16	0.21	3.09	0.18	2.67	0.20	4.23
age 17-19	-0.09	-1.10	-0.07	-0.86	-0.08	-1.47
province dummies	Yes		Yes		Yes	
time dummies	Yes		Yes		Yes	
		SE		SE		SE
$\rho$	3.00	0.16	3.49	0.20	3.20	0.12
$\theta$	0.40	0.12	0.45	0.13	0.43	0.08
$N$	3595		4007		7602	
<i>Log likelihood</i>	-685.8		-616.1		-1324.6	

Note:  $\rho$  is the baseline hazard shape parameter (increasing when  $\rho > 1$ , as here). $\theta$  is the frailty or unobserved heterogeneity variance (Lancaster, 1979).

The girls and boys regressions are allowed to be pooled:

$$\text{chi2}(36) = 45.26; \text{Prob} > \text{chi2} = 0.14$$

**Table 4. Weibull AFT Model with Heterogeneity (Clustered Income)**

Dependent variable: schooling duration

	Girls		Boys		All	
	Coef.	z	Coef.	z	Coef.	z
water	-0.22	-4.13	-0.06	-1.37	-0.05	-1.15
female					0.03	1.12
water × female					-0.16	-2.49
clustered income	0.29	5.71	0.23	5.47	0.26	7.92
children's household work	-0.06	-4.39	-0.07	-2.27	-0.06	-5.29
children's market work	-0.12	-7.50	-0.09	-8.14	-0.11	-10.83
single child (ref)						
one older brother	-0.06	-1.10	-0.10	-1.89	-0.09	-2.30
one older sister	0.08	1.02	-0.07	-1.40	-0.03	-0.58
one younger brother	-0.06	-0.73	-0.19	-3.04	-0.13	-2.66
one younger sister	0.01	0.07	0.02	0.27	0.02	0.34
two/more siblings	-0.06	-1.09	-0.03	-0.68	-0.05	-1.47
age 6-11 (ref)						
age 12-16	0.16	2.31	0.13	1.84	0.16	3.13
age 17-19	-0.17	-1.98	-0.15	-1.76	-0.16	-2.64
province dummies	Yes		Yes		Yes	
time dummies	Yes		Yes		Yes	
		SE		SE		Se
$\rho$	2.87	0.16	3.27	0.19	3.06	0.12
$\theta$	0.48	0.12	0.38	0.13	0.45	0.09
$N$	3604		4013		7617	
<i>Log Likelihood</i>	-722.5		-652.1		-1393.6	

**Notes:** The high and low income categories are based on k-means clustering using per capita household income, parental job status and parental educational qualifications. The cut-off point of per capita household income is set at its median value, while job status and educational qualifications variables are kept with their four original categories. We classify the data into two clusters “poor” (9572 subjects) and “rich” (2962 subjects).

$\rho$  is the baseline hazard shape parameter (increasing when  $\rho > 1$ , as here).

$\theta$  is the frailty or unobserved heterogeneity variance (Lancaster, 1979).

The girls and boys regressions are allowed to be pooled:

$$\text{chi2}(24) = 32.86; \text{Prob} > \text{chi2} = 0.11$$

**Table 5. Weibull AFT Model with Heterogeneity (Rich vs. Poor Girls)**

Dependent variable: schooling duration of girls

	Poor Girls		Rich Girls	
	Coef	z	Coef	z
water	-0.20	-3.51	-0.30	-2.12
children's household work	-0.06	-3.91	-0.07	-2.37
children's market work	-0.12	-7.17	-0.11	-1.62
single child (ref)				
one older brother	0.01	0.19	-0.35	-2.78
one older sister	0.10	1.20	0.02	0.12
one younger brother	-0.10	-1.15	0.29	1.03
one younger sister	0.05	0.49	-0.14	-0.61
two/more siblings	-0.05	-0.81	-0.03	-0.16
age 6-11 (ref)				
age 12-16	0.15	1.88	0.25	1.62
age 17-19	-0.18	-1.88	-0.14	-0.79
		SE		SE
$\rho$	2.87	0.17	2.93	0.43
$\theta$	0.46	0.13	0.54	0.35
$N$	2637		967	
<i>Log Likelihood</i>	-564.2		-142.2	

Note: The rich and poor categories are set on the basis of the clustered income, for details see notes under the table 5. The impact of poor access to water is not found to be statistically different in these two regressions:  $\chi^2(1) = 0.44$ ;  $\text{Prob} > \chi^2 = 0.51$ .

**Table 6. Probability of School Enrolment (Marginal Effects) Girls only**

	ME	z	ME	z
tap water - home (water1) (ref)				
tap water - courtyard (water2)	0.06	1.48	0.04	1.01
well water - courtyard (water3)	0.02	0.45	-0.01	-0.38
other water outside courtyard (water 4)	0.02	0.40	-0.03	-0.60
water2 × period	-0.05	-0.73	-0.06	-1.00
water3 × period	-0.07	-1.44	-0.09	-1.74
water4 × period	-0.12	-1.93	-0.13	-2.05
period	-0.08	-1.77	-0.07	-1.46
log income per capita	0.02	1.60		
father - high occ. status (ref)				
father - medium occ. status	-0.09	-2.06		
father - no wage work/farmer	-0.16	-5.38		
father - other/unemployed	-0.26	-3.26		
mother - high occ. status	0.13	2.28		
mother - medium occ. status (ref)				
mother - no wage work/farmer	-0.06	-1.54		
mother - other/unemployed	-0.09	-1.85		
father - no school qualification	-0.15	-3.75		
father - primary school	-0.06	-1.91		
father - juniour middle school	-0.02	-0.84		
father - high school/college (ref)				
mother - no school qualification	-0.04	-1.08		
mother - primary school	-0.01	-0.39		
mother - juniour middle school	0.01	0.18		
mother - high school/college (ref)				
clustered income			0.12	6.09
children's household work	-0.04	-4.58	-0.04	-4.89
children's market work	-0.13	-5.43	-0.14	-5.28
single child (ref)				
one older brother	-0.01	-0.20	0.01	0.29
one older sister	0.01	0.40	0.02	0.44
one younger brother	-0.02	-0.73	-0.02	-0.65
one younger sister	0.02	0.69	0.02	0.64
two/more siblings	-0.02	-0.57	-0.01	-0.28
age	0.30	15.53	0.30	15.15
age squared	-0.01	-16.84	-0.01	-16.61
county (36) and wave (6) dummies	yes		yes	
Number of observations	4870		4879	
Number of groups	2297		2300	

Note: marginal effects are calculated at the means of the independent variable. The generalised estimating equations model is implemented using STATA xtgee programme with a specification of binomial family; logit link function and exchangeable correlation structure. If there is a balanced panel, this specification is equivalent to a population averaged panel logit model (Hardin and Hilbe, 2003).

**Table 7. Relative School Attainment (Marginal Effects using XTGEE)**

	ME	z	ME	z	ME	z
tap water - home (water1) (ref)						
tap water - courtyard (water2)	0.04	0.60	0.02	0.29	0.03	0.34
well water - courtyard (water3)	0.02	0.29	-0.02	-0.26	-0.01	-0.13
other water outside courtyard (water4)	-0.06	-0.81	-0.09	-1.15	-0.10	-1.24
water 2 × period	-0.14	-1.48	-0.11	-1.18	-0.13	-1.31
water 3 × period	-0.20	-2.38	-0.13	-1.63	-0.15	-1.80
water 4 × period	-0.41	-3.99	-0.34	-3.30	-0.32	-3.11
period	0.24	2.83	0.20	2.37	0.21	2.51
relative clustered income	0.48	8.09	0.10	0.93	0.22	1.70
water 2 × relative clust. income			0.32	2.08	0.10	0.45
water 3 × relative clust. income			-0.16	-1.20	-0.27	-1.44
water 4 × relative clust. income			0.36	1.99	0.01	0.06
relative clust. income × period			0.67	6.17	0.46	2.70
water 2 × relative clust. income × period					0.41	1.33
water 3 × relative clust. income × period					0.17	0.65
water 4 × relative clust. income × period					0.82	2.31
relative household work	-0.10	-4.67	-0.10	-4.60	-0.10	-4.60
relative market work	-0.14	-9.57	-0.14	-9.41	-0.14	-9.45
age dummies (age 6-19)	yes		yes		yes	
county dummies (36 dummies)	yes		yes		yes	
wave dummies	yes		yes		yes	
Number of observations	4809		4809		4809	
Number of groups	2095		2095		2095	

Note: marginal effects are calculated at the means of the independent variable. The generalised estimating equations model is implemented using STATA xtgee programme with a specification of gaussian family; identity link function and exchangeable correlation structure. If there is a balanced panel, this specification is equivalent to a random effects panel model (Hardin and Hilbe, 2003).

**Table 8. Percentile of survival distributions using Weibull – various scenarios**

		<i>Survival Time</i>	0.95	0.75	0.50	0.35	0.25
Good water (1)	Girls		4.5	8.2	11.2		
	Boys		4.5	7.9	10.6		
<b>Poor access to water (2)</b>	<b>Girls</b>		<b>3.5</b>	<b>6.5</b>	<b>9.2</b>	<b>10.8</b>	<b>12.0</b>
	Boys		3.7	7.2	9.9	12.0	
High income (upper cluster) (3)	Girls		5.4	9.5	12.0		
	Boys		5.5	9.4	12.0		
Low income (lower cluster) (4)	Girls		3.8	6.9	9.7	12.0	
	Boys		4.6	7.4	9.9	11.6	
Good water High income (5)	Girls		5.3	9.7	12.0		
	Boys		5.5	9.5	12.0		
<b>Poor access to water High income(6)</b>	<b>Girls</b>		<b>4.2</b>	<b>7.8</b>	<b>10.9</b>		
	Boys		5.4	8.8	12.0		
Good water Low income (7)	Girls		3.9	7.3	10.2	12.0	
	Boys		4.5	7.5	9.9	11.7	
<b>Poor access to water Low income (8)</b>	<b>Girls</b>		<b>3.3</b>	<b>5.7</b>	<b>8.3</b>	<b>9.8</b>	<b>11.3</b>
	Boys		4.3	6.9	9.4	11.0	12.0

**Notes:** Weibull AFT with heterogeneity specification used to create the distributions; variables specified for a typical value in model (1) – (8) are listed in the first column. Other variables in the model are set to their mean values. Poor access to water here includes only category 4 (water outside the courtyard).

**Table 9. Determinants of Community Average School Enrolment Rate (OLS)**

	Girls		Boys	
	Coef.	t	Coef.	t
community rate of water 4 (other water outside courtyard) (1)	0.08	1.29	0.00	-0.01
community rate of girls' periods (2)	-0.18	-3.70		
(1) × (2)	-0.34	-2.80		
community proportion of rich people	0.08	3.10	0.08	3.00
community average household work (girls or boys)	-0.02	-1.33	0.01	0.23
community average market work (girls or boys)	-0.09	-9.17	-0.11	-9.52
community average number of children	-0.03	-1.41	0.01	0.44
community average age of survey individuals	-0.01	-1.32	-0.03	-5.25
community dummies (136)	yes		yes	
wave dummies (6)	yes		yes	
Number of Observations	526		524	
Adj R squared	0.46		0.36	

Note: OLS is permitted in all the regressions, so the specification is a simple pooled .

## References

- Barro, R and Lee J, 2000, "International Data on Educational Attainment: Updates and Implications" (CID Working Paper No. 42, April 2000), Center for International Development at Harvard University Research Datasets
- Becker, G, 1981, A Treatise on the family, Cambridge: Harvard University Press
- Becker, G, & Lewis, G, 1976, "On the interaction between the quantity and quality of children", The Economic Approach to Human Behaviour, University of Chicago Press, Chicago and London
- Becker, G., S., 1973, "A theory of marriage: part 1", *Journal of Political Economy*, 18(4): 813-46
- Beek J.S. et al., 1996, "Puberty and Dysmenorrhea Treatment", Novice's Gynecology, Williams and Wilkins Publication, Inc. 771-780.
- Behrman, J., R., 1996, "Impact of health and nutrition on education", *World Bank Research Observer*, 11(1): 23-37
- Behrman, J., R., Khan, S., Ross, D. & Sabot, R., 1997, "School Quality and Cognitive Achievement Production: A Case Study for Rural Pakistan", *Economics of Education Review*, 16 (2): 127-142
- Behrman, J., R., Foster, A., D., Rosenzweig, M., R. & Vashishta, P., 1999, "Women's Schooling, Home Teaching and Economic Growth", *Journal of Political Economy*, 107(4): 682-714
- Benjamin, D, Brandt, L, Giles, J, Wang, S, 2005, "Income Inequality During China's Economic Transition", working paper, <http://www.chass.utoronto.ca/~benjamin/BBGW.pdf>
- Bista M B, 2004, "A Review of Research Literature on Girls' Education in Nepal", Report prepared for UNESCO Bangkok, <http://unesdoc.unesco.org/images/0013/001386/138640e.pdf>
- Brown, P, & Park, A, 2002, "Education and poverty in rural China", *Economics of Education Review*, 21: 523-541
- Burrows, G, J Acton and T Maunder, 2004, Water and Sanitation: the Education Drain, London: Wateraid, <http://www.wateraid.org/other/startdownload.asp?DocumentID=113>
- Cairncross, S. et al., 1998, "The public and domestic domains in the transmission of disease", *Tropical Medicine and International Health*, 1: 27-34.
- Cameron, A. C., Trivedi, P. K., 2005, Microeconometrics – Methods and Applications. Cambridge University Press
- CHNS, 2007. China Health and Nutrition Survey, Chapel Hill, North Carolina: Carolina Population Center, <http://www.cpc.unc.edu/projects/china>
- Colclough, C., Rose, P., Tembon, M., 2000, "Gender inequalities in primary schooling, the role of poverty and adverse cultural practice", *International Journal of Educational Development*, 20: 5-27
- Conelly, R, & Zheng, Z, 2003, "Determinants of school enrolment and completion of 10 to 18 year olds in China", *Economics of education review*, 22: 379-388.
- Cox, D.R., & Oakes, D., 1984, Analysis of Survival Data. London: Chapman Hall
- Davison, J., Kanyuka, M., 1992, "Girls participation in basic education in southern Malawi", *Comparative Education Review*, 36(4): 446-466
- Dagwood M.Y., 1995, "Dysmenorrhea", *Journal of Reproductive Medicine*, 30(3): 154-167
- Drank Shayani D.K. and Uenkafa R.P., 1995, "A Study on Menstrual hygiene Among Rural Adolescent Girls", *Indian Journal of Medical Sciences*, 45: 139-143.

- El-Gilany, A. H., Badawi K., El-Fedawyb, S, 2005, “Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt”, *Reproductive Health Matters*, 13(26): 147-152
- Emerson, P.M., Souza, A.P., 2007, Child Labor, School Attendance, and Intrahousehold Gender Bias in Brazil, *The World Bank Economic Review Advance Access published March 8, 2007*, <http://wber.oxfordjournals.org/cgi/reprint/lhm001v1>
- Foster., E.,M., 2002, “How Economists Think About Family Resources and Child Development”, *Child development*, 73(6), 1904-14
- Ghisletta P and D Spini, 2004, “An introduction to generalised estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals”, *Journal of Educational and Behavioural Statistics*, 29 (4): 421-437.
- Hannum, E, 1998, “Educational inequality: hidden consequences of the reform era in rural China: 1949-1985. *Research in Social Stratification and Mobility*, 13, 73-98
- Hannum, E, 2005, “Market transition, educational disparities, and family strategies in rural China: new evidence on gender stratification and development”, *Demography*, 42(2): 275-298
- Hannum E and C Buchmann, 2005, “Global education expansion and socio-economic development: An assessment of findings from the social sciences”, *World development*, 33 (3); 333-354.
- Hardin, J., W. & Hible, J., M., 2003, *Generalized Estimating Equations*, New York: Chapman & Hall/CRC
- Hill, M., A., King, E., M., 1995, “Women’s education and economic well-being”, *Feminist Economics*, 1(2): 21-46
- Hill, M., A., King, E., M., 1993, “Women’s education in developing countries: An overview”, *Women’s education in developing countries*, The John Hopkins University Press, Baltimore and London
- Holmes, J., 2003, “Measuring the determinants of school completion in Pakistan, Analysis of censoring and selection bias”, *Education of Economics Review*, 22, 249-264
- Hosmer, D, & Lemeshow, S, 1999, *Applied Survival Analysis*, New York: Wiley.
- Kalbfleisch, J.D., Prentice R.L., 1980, *The Statistical analysis of failure time data*, Wiley and Sons, New York.
- Kirk, J. and Sommer, M. 2006, “Menstruation and body awareness: critical issues for girls’ education”, *EQUALS*, 15: pp 4-5.
- Knight, J, & Li, S, 1996, “Educational attainment and the rural-urban divide in China”, *Oxford Bulletin of Economics and Statistics*, 58: 83-117
- Li, D, & Tsang, M.C, 2003, “Household education decisions and implications for gender inequality in education in rural China”, *China: An International Journal*, 2003, 1(2): 224-248.
- Lidonde, R., 2005, “Why gender is important in SSHE”, *WaterLines*, 23(3), 6-9
- Nahar, Q, 2006, “Addressing special needs of girls – challenges in school”, Second South Asian Conference on Sanitation, 20-21 September, Islamabad, Pakistan
- Poureslami M, Osati-Ashtiani F, 2002, “Attitudes of female adolescents about dysmenorrhea and menstrual hygiene in Tehran suburbs”, *Archives of Iranian Medicine*, 5(4): 219-224
- Schultz, T, 2002, “Why governments should invest more to educate girls”, *World Development*, 30 (2): 207-225.

- Secerino SK and Moline M.E., 1995, "Premenstrual Syndrome Identification and Management". *Drugs*. 71-82.
- Singh MM, et al., 1999, Awareness and health seeking behaviour of rural adolescent school girls on menstrual and reproductive health problems, *Indian Journal of Medical Science*, 53(10):p 439-43
- Snel, M, & Shordt, K, 2005, "The evidence to support hygiene, sanitation and water in schools", *WaterLines*, 23(3), 2-5
- Song, L., & Appleton, S., & Knight, J. , 2006, "Why do girls in Rural China have lower school enrolment?" *World Development*, 34(9), 1639\_1653
- Tharoor, S. 2007, *The Elephant, the Tiger and the Cell Phone: India—The Emerging 21st Century Power*, New York: Arcade Publishing.
- Tsui, M, & Rich, L, 2002, " The only child and educational opportunity for girls in urban China", *Gender and Society*, 16(1): 74-92
- UNICEF, 2008, "Water, environment and sanitation"  
[[http://www.unicef.org/wes/index\\_womenandgirls.html](http://www.unicef.org/wes/index_womenandgirls.html)]
- UN, 2006, Millennium Development Goals Indicators Database. New York: United Nations, Department of Economic and Social Affairs, Statistics Division  
[<http://mdgs.un.org>]
- Yang J H, 2006, "Has the One-Child Policy Improved Adolescents' Educational Wellbeing in China?", CHNS discussion paper 2006.
- Yueh, L, 2006, "Parental investment in children's human capital in urban China", *Applied Economics*, 38: 2089\_2111