北京大学中国经济研究中心
China Center for Economic Research
讨论稿系列
Working Paper Series

No. E2006011 2006-09

Health Shocks and Children's School Attainment in Rural China

Ang Sun & Yang Yao

NO. E2006011 Oct. 9, 2006

Health Shocks and Children's School Attainment in Rural China

Ang Sun & Yang Yao

NO. E2006011 Oct. 9, 2006

Health Shocks and Children's School Attainment in Rural China[*]

Ang Sun
Graduate student
Department of Economics
Brown University
&
Yang Yao
Professor
China Center for Economic Research
Peking University

Using a long panel dataset of farm households from China covering the period of 1987 to 2002, this paper studies how major health shocks happening in a family affect its children's school attainment. We find that primary school-age children are the most vulnerable to health shocks. Their chances to enter middle school decrease by 19.4% when a prime-age adult in their families has a major illness. Our robustness tests of using a more homogenous sub-sample, instrumenting the health shocks, estimating a family fixed-effect model, and controlling the sibling effect have found smaller but still statistically and economically significant effects. We have also found weak evidence for the screening effect of health shocks. Policy implications of these findings are briefly discussed.

***JEL classification*:** I2; D1; E2
Keywords: school attainment; health shocks; human capital

---

[*] A previous version of this paper was presented at the Chinese Economists Society 2005 annual conference in Chongqing and the 2005 China Economics Annual Conference in Xiamen. We are grateful for the participants of those two conferences as well as Anna Aizer, Josh Angrist, Daniel Bromley, Keith Maskus, Gerard Pedro I Miquel, Nancy Qian, and John Strauss for their comments. The Research Center of Rural Economy, the Ministry of Agriculture, the People's Republic of China provided assistance in data collection. Financial supports from the Chinese Medical Board, the World Bank, and the National 211 Project are greatly appreciated. Corresponding author: Yang Yao, China Center for Economic Research, Peking University, Beijing 100871, China. Phone: 86-10-6275-3103; fax: 86-10-6275-1474; email: yyao@ccer.pku.edu.cn.

1. Introduction

Major illnesses are the most unpredictable and devastating shocks for farm households in developing economies because most of them do not have adequate health insurance. A family loses on two fronts when a major health shock comes. On the one hand, it has to spend a considerable amount of money to treat the sick member; on the other hand, it loses part of its labor capabilities when the sick member is a major laborer in the household (Gertler and Gruber, 2002). These two factors have both short-term and long-term effects on the family. In the short run, the family has to reduce its consumption due to imperfect insurance (Townsend, 1994; Gan, Xu, and Yao, 2006a); in the long run, the family may fail to accumulate enough productive assets including children's education. As a result, experiencing a major health shock may well send a family to a prolonged poverty trap (Gan, Xu, and Yao, 2006b).

The study of natural shocks' long-term impacts on children's educational attainment is difficult without data with sufficiently long time coverage. As a result, most studies concentrate on their impacts on households' short-term educational investment and children's school enrollment (e.g., Rosenzweig and Wolpin, 1993; Jacoby and Skoufias, 1996; Thomas et al., 2004). However, the impacts of the shocks may be temporary because a child may go back to school again after the family passes the period of several liquidity constraints. Studies based on short panel data cannot provide a definitive assessment to this effect.

Using a unique dataset collected from rural China with a total of 1,144 households for the time period of 1987 to 2002, this paper provides a study on how major health shocks affect children's school attainment in a developing economy. The long coverage of the data allows us to capture a large number of health shocks. In our sample period, about 60% of the households experienced at least one health shock, and 44.9% of them experienced at least one health shock on their prime-age adult members. We adopt the sequential response model to study how a health shock on an adult family member affects children's probability of success in each stage of education up to high school entrance. In addition to finding health shocks' long-term impacts on school attainment, we will also study the issue whether some stage (or stages) is more sensitive to health shocks than others in terms of people's dropout rates. Jacoby and Skoufias (1996) argue that the demand for labor is a direct reason for families to stop their children's education in response to a major illness. This implies that older children are more likely to stop education than younger ones. However, Thomas et al. (2004) find in Indonesia that the Asian Financial Crisis took its greatest toll on poor children who had older siblings.

Our results are more consistent with the finding of Thomas et al. (2004). We find that a child's probability to finish primary school drops by 12.3% when a prime-age adult in his family has a major illness in his primary school age. Conditional on graduation from primary school, the shock further reduces his chances to enter middle school by 8.1%. This means that the overall effect of the shock is to reduce a child's probability to enter middle school by 19.4%. The actual dropout rate up to the stage of middle school entrance was 22.8% for children whose families experienced a health shock in their primary school age. The effect of a health shock is equivalent to 80.5% of it. However, health shocks happening in one's primary school age do not have any significant impact on his probability to finish middle school or to enter high school. On the other hand, shocks happening before one's primary school age increase one's probability to finish middle school but do not affect his probability to finish primary school or to enter middle school. These two results show that health shocks have the screening effect in the sense that they select out families that have higher expectations on their children's education or children with higher academic potentials than ordinary families or children. If a child from a shock-hit family finishes one stage of education and moves on to the next stage of education, there is a large chance that either his family has high expectation on him, or he has demonstrated higher academic potentials than other kids.

A major health shock is defined in our paper as an illness that required hospitalization or expenditure larger than 5000 yuan. This definition may truncate the report of illness by household income. To check the robustness of our results, we have done several sensitivity tests including using a more homogenous sub-sample, instrumenting health shocks, and taking care of household fixed effects. These tests obtain smaller estimates for the impacts of the health shocks, but they are still statistically and economically significant.

Another problem with our baseline results is that we do not have enough control on the sibling effect. This effect can arise for two reasons. One is related to the order of children in obtaining education within a family. It is possible that children who start earlier in their education have the priority to finish school first. The second is related to the relative costs of education for children at different stages of education. Although Jacoby and Skoufias (1996) hypothesize that older children are more likely to be hurt by a shock, the other side of the story may also hold. Older children are closer to graduation than younger children, so it is relatively cheaper for a family to let them finish their education than letting younger children to continue. To the extent that young children are more likely to have older siblings, our finding that young children are more vulnerable to family health shocks may be a result of the sibling effect rather than one of health shocks themselves. Although we have included in our regressions variables describing a child's siblings, they may not be enough to deal with this problem. We test our baseline results against this problem by including in the sample only the first child experiencing a shock so we do not need to compare the older and younger children within the same family that was hit by a health shock. The effects of health shocks are indeed found to be smaller, but they nevertheless remain statistically significant.

The rest of the paper is organized as follows. Section 2 introduces the data and presents some descriptive results. Section 3 presents the baseline results. Section 4 conducts the robustness tests on the baseline results. Section 5 concludes the paper.

2. Data and descriptive evidence

2.1 Data

Our data come from two sources. One is the National Fixed-point Survey (NFS) maintained by the Research Center of Rural Economy, the Ministry of Agriculture, People's Republic of China, and the other is a retrospective survey conducted by the authors in the spring of 2003. The NFS is a longitudinal survey of about 24,000 households of more than 340 villages in all continental Chinese provinces. It started in 1986 and provides rich information on household income and consumption as well as village characteristics. The stratified random sampling strategy was adopted when the survey was first started. Despite attritions, it has generally maintained a panel structure although mis-numbering of households exists. Our retrospective survey was conducted on 48 villages of 8 provinces, Guangdong, Zhejiang, Hunan, Henan, Shanxi, Jilin, Sichuan, and Gansu. The major purpose of the survey was to obtain information on family history of major illnesses in the period of 1987-2002. It also collected data on the educational attainment of each family member by 2002. To ensure a panel structure for the data, we used several combined criteria based on household characteristics (the size of the house, land area, number of people, and the age of the head) to identify and match households in the NFS sample. Consequently, in the eight provinces covered by our survey, 1,144 households and 8,414 persons were left in the sample. The exclusion is likely to be random as no systematic attrition and change of households were reported in the NFS. As a result, our samples can be treated as a random sample.

2.2 Health shocks

We define a major health shock as an illness that requires hospitalization or a total spending over 5000 RMB. We adopted this definition for practical reasons. The period covered by our survey is very long, so we had to restrict our attention to major health shocks so as to ensure the accuracy of people's recounting of their families' health history. In this paper, we will only study the major shocks on family members between age 16 and 60, that is, the shocks that were received by family labors. When a family labor has a serious illness, the family needs to both spend money and forgo current and future income, so a shock on a family labor is likely to have the most serious impacts on children's school attainment.

[Figure 1 about here]

Our definition of a major health shock has the potential to cause a truncation on the report of illness along the line of income. It is quite possible that some households experienced a major health shock but nevertheless did not spend more than 5000 RMB or send the patient to hospital because they did not have enough money. This could be a problem for our estimation as our data covers a long period of 16 years and income experienced significant growth in the period.[1] This is evident in Figure 1 that shows the number of major illnesses reported for each year. While there was no significant trend before 1994, the number of reports increased quickly after that year although by 1996 the number was still in the range of those in the early years.[2]

In our baseline regressions, we will handle the possible report biases by controlling the year of school entrance and per-capita household income of 1987. The year of school entrance controls temporary factors that are common to all the children with roughly the same school age, including the effects brought about by the growth of income. Household income controls persistent family characteristics that could lead to biased illness reports. Of course, the control cannot be effective for shocks happening in later years. We will improve upon it by several other more sophisticated methods when we conduct the robustness tests.

---

[1] Income growth in our sample villages was substantial over the period of 1987-2002. The average per-capita income in 2002 was 2.2 times of that in 1987 in real term, which implies an average annual growth rate of 5.4%.

[2] Memory loss can also be a reason for the pattern shown in Figure 2. However, this explanation cannot explain why the number of reports did not have any trend before 1994 but began to increase steadily after that year.

2.3 School attainments in the whole sample

Before conducting detailed analysis based on different stages of education, we first look at the school attainment in the whole sample. To capture a person's entire pre-college schooling history, we study the sample of people between 19 and 32 years old in 2002. The youngest of this cohort would have finished high school by 2001, and the oldest would have done so in 1988. School attainments among this sub-sample were not impressive. Only 87.0% of them finished primary school, 51.4% finished middle school, and 19.7% entered high school. The average school attainment was 8.4 years in school. Figure 2 shows the average final schooling years of people who entered primary school in different years. It is apparent that younger people had higher school attainment. In 1986, China started to make it compulsory for children to get middle school education (that is, a total of nine years in school). There was indeed a small jump of school attainment for people who entered primary school in that year. In addition, the average schooling years have been closing to 9 years starting with the people who were supposed to graduate from middle school in the early 1990s.

[Figure 2 about here]

Figure 3 provides a comparison between people whose families received at least one health shock in 1987-2002 and those not by controlling per-capita family income of 1987, father's education, and mother's education, respectively. The lines are moving-average trends of actual data. It is evident that school attainment increased with income and parents' education in both groups, but people in the shock-hit group had consistently lower attainments than those in the shock-free group. It is particularly noteworthy that the gaps were larger for higher income and parents' education. One possible explanation is that on the one hand, education attainment of the shock-free group was already low (lower than the 9 years required by the government) when income or parents' education was low, and on the other hand, families with higher income or families with well-educated parents tended to have higher expectations for children's education and a health shock cut this plan more severely.

[Figure 3 about here]

2.4 Age cohorts for analysis

There are two ways to characterize educational decision (Cameron and Heckman, 1998). One is to treat it as deterministic and use a linear model to predict a person's school achievement or decision to invest in education, and the other is to treat it as a stochastic process that is subject to random shocks in each stage of education. In the context of a rural and poor economy, the second approach is a more sensible one. It is hard for a family to make a deterministic plan for its children's educational achievements because there are many uninsured shocks along the way of the children's education. This is especially pertinent to health shocks studied by this paper because illnesses are the most unpredictable idiosyncratic risk. The sequential response model is the most suitable model for our purposes. To use this model, we divide pre-college education into four stages: to finish primary school, to enter middle school, to finish middle school, and to enter high school. The stage of entering primary school is ignored because only a negligible number of people did not have any education. Among the 1,489 people in the 2002 age cohort of 13 – 29 which our econometric analysis will primarily work with, only 0.7% did not have any education. The percentage of high school graduation among the cohort was 10.8%. So there would be only 161 observations if we were to study this stage of education. So we ignore it.

Apparently, a child cannot enter the next stage without finishing the current stage of education. Next we will define the sample of cohort that we will study for each stage. Before we begin this task, however, we need to consider two factors regarding a health shock's impacts on people's educational decision. The first is that it may take some time for the impacts to realize. If a family received a shock in the middle of a semester, it is unlikely that this family would immediately stop its children's education because it had already paid the tuition and other fees for this semester. The second is that health shocks' impacts may be temporary and people may drop back in after a semester or two. Therefore, we will allow one year for health shocks to show their impacts and one year for people to drop back in. This means that we will study people's educational decisions between 1988 and 2001 (inclusive). One option to define the cohort for the study of a stage of education is to only include people whose entire spell of education in that stage is observed. For example, to study the decision to finish primary school, we may only include people who were supposed to enter primary school in or after 1988 and finish primary school in or before 2001. But this option will lose people who entered primary school before 1988 but graduated after that. To maximize the benefits provided by our data, we will include people whose potential graduation year from a specific stage falls into 1988-2001.[3]

We need to define a separate cohort for each stage of education. It turns out that they are the primary school cohort and the middle school cohort and their sub-samples. The primary school cohort is comprised of people who were supposed to gradate from primary school between 1988 and 2001 (inclusive), and the middle school cohort is comprised of people who had finished primary school and are supposed to graduate from middle school in the same period. Taking age 6 as the year to start school, we then include in the primary school cohort people between 13 and 26 years old (inclusive) in 2002 who actually entered primary school at some point of their life. People of age 13 would have finished primary school by 2001 (so they would have at least one year of reconsideration if they dropped out), and people of age 26 would have done so in 1988 (so shocks happening in 1987 would show their impacts). As a check to our results, we will also try the cohort of age 13 – 21 who would have their entire primary school education in the period of 1988-2001. The middle school cohort is defined in a similar way. The age of entering middle school is usually 13 years old and the graduation age is usually 15 years old. So we define the cohort as the group of people who were between 16 and 29 years old (inclusive) in 2002 and who actually entered middle school at some point of their life. The youngest of this cohort would finish middle school in 2001, and the oldest would do so in 1988. As in the case of the primary school cohort, we will also try the cohort of age 16 – 26 who would have their entire middle school education in the period of 1988-2001.

2.5 The role of health shocks on educational attainments

There are a total of 1,489 people between 13 and 29 years old, the union of the primary and middle school cohorts. Among them, 254 (17.1%) experienced one or more health shocks in their families before they finished middle school, and the rest did not experience any. The gap between the school attainments of the two groups of people was not significant, though, because among the shock-hit group, 88.2% finished primary school, 77.6% entered middle school, 46.9% finished middle school, and 18.1% entered high school while the corresponding percentages for the shock-free group were 87.4%, 77.8%, 47.4%, and 18.7%. However, a close study based on the cohorts reveals more interesting results.

[3] The number of people whose families' health history is not observed is not large, though. For people who graduated in 1988, we have one year (1987) to observe their families' health history; for people who graduated in 1989, we have two years of observations, and so on.

[Table 1 about here]

Table 1 compares the shock-hit and shock-free groups based on the primary and middle school cohorts. Panel A studies the primary school cohort. Comparisons are being made based on the timing of the shock, that is, between people who received a shock before they entered primary school (that is, before they turned 6 years old) and those who did not, and between people who received a shock in primary school (that is, when they were between 6 and 12 years old) and those who did not, respectively. It is evident that shocks happening in and before primary school had quite different impacts on school attainments. Receiving a shock in primary school had a negative impact on one's chances to finish primary school and to enter middle school. In contrast, people receiving a shock before primary school had less than half of the dropout rate of those without although their chances to enter middle school are still lower. While the negative impacts of the shocks fall in line with the conventional wisdom, the positive impact of the shocks before primary school needs careful interpretation. It seems that it reflects the screening effect of health shocks. However, almost all the people in our sample entered primary school so it is unlikely that screening happened unless health shocks only dropped on families with higher expectations for their children's education.[4] The positive effect may reflect the composition of people who received a shock before their primary school age and will vanish when personal, family, and village characteristics are controlled in econometric analysis.

However, the screening effect may indeed exist in the middle school cohort with respect to school dropouts. This is evident in Panel B of Table 1 where the middle school cohort is studied. People who experienced a shock before or in their primary school age were less likely to drop out than those without and the gaps were both around 10 percentage points. Since people were tested by their primary school study, we are not sure whether the selection was based on higher expectations or better academic potentials. In terms of the chances to enter high school, having a shock before primary school had a significant negative effect, but having a shock in primary school had a positive effect. However, the number of observations is small in both cases so these two results may not have statistical significance.

Shocks happening in a person's middle school age had mixed effects. Compared

[4] Because chances are small for a family to know the academic potential of its children before they enter primary school, the screening effect would solely come from a family's higher expectation for its children's educational achievements if it existed.

with people who did not receive a shock in their middle school age, people hit by a shock were less likely to enter high school, but were slightly more likely to finish middle school.

It is noteworthy that the comparisons in Table 1 do not reveal the marginal effects of the health shocks at different stages of education because they are made between people receiving a shock in a specific stage of education and those who did not in that stage, but the marginal effect should be based on the comparison between the former group and those who did not receive any shock in any stage of education. However, the results are still indicative by offering some important aspects for our econometric analysis to explore.

3. Baseline results

Consistent with the four stages of education, we will study four dichotomous decision variables, to finish primary school, to enter middle school, to finish middle school, and to enter high school, each using a cross-sectional sample derived from the primary school and middle school cohorts. The baseline model that we will estimate is the following probit model:

$$(1) \quad y_{ijk} = \alpha_0 + S_{ijk}\beta + X_{ijk}\gamma_1 + X_{ij}\gamma_2 + X_i\gamma_3 + \varepsilon_{ijk}\,.$$

In the equation, y_{ijk} is the notional decision variable for the kth child of family j in village i to finish a specific stage of education. The child finished a specific stage of education by 2002 if y_{ijk} is greater than zero; otherwise, he did not. S_{ijk} is a set of dummy variables each of which indicates whether a child experienced a shock in a specific stage of education. They are our major concerns. The match between S_{ijk} and y_{ijk} will be discussed later when we study each cohort. X_{ijk}, X_{ij}, and X_i are sets of control variables at the personal, family, and village levels, respectively. The error term ε_{ijk} is assumed to be i.i.d. although we will use the heteroscedasticity-corrected estimator to estimate the standard errors. Lastly, α_0, β, and the γ's are parameters to be estimated.

3.1 Control variables

Our control variables are selected based on the discussions in the literature (e.g., Schultz 1999). Among the personal variables, we have included gender, a set of variables describing the number of siblings, and two sets of dummy variables indicating the year when a person began primary school and middle school, respectively (thereafter called the school entrance dummies. The first enters the regressions based on the primary school cohort and the second enters the regressions based on the middle school cohort). While the inclusion of gender in the regressions is self-explaining, we offer some discussions to the other variables. The school entrance dummies control the time factors that had common effects on all the people at the same particular stage of education. They also substitute for a person's age as the year of starting school is linearly correlated with age. The set of variables for siblings is meant to control intra-family competition and complementarities for resource allocations. Cameron and Heckman (1998) only use the number of siblings in their study. We adopt a finer representation to include four variables in our regressions, indicating the number of older brothers, older sisters, younger brothers, and younger sisters, respectively. While more siblings imply more competition for resources so all the four variables would have negative impacts on one's school attainment, the opposite may also be true for having an older brother or sister. It is possible that when it was hit by an unexpected shock, a family would let older children drop out because they could immediately enter the labor market (Jacobs and Skoufias, 1996), so younger children could continue their schooling. In addition, having an older sister could have an extra advantage in the Chinese context because the traditional culture expects the elder daughter to sacrifice for her younger siblings when harsh times come.

Among the family variables, we have included per-capita household net income in 1987, father's schooling years, mother's schooling years, and father's occupation in 1987. The 1987 income is meant to capture a family's financial preparation for its children's education. This could be problematic for later years as the coverage of our data is long. We will address this issue by studying the sub-sample of years before 1996 in the next section when we present our robustness tests. Parents' schooling years are standard control variables in the literature. It is usually found that mother's education is more important than father's for children's school attainment (e.g., Shultz 1999 and a recent study by Brown 2006 on rural China), but exceptions also exist (e.g., Li, Lin, and Yao 2002 on rural China). Father's occupation is a dummy variable indicating whether the father was not a full-time farmer (so it is equal to zero if he was a full-time farmer and equal to one if he was not). This variable is meant to control a family's expectations for its children's education. It is natural to expect that full-time farmers have lower expectations for their children's education.

Lastly, the set of village variables control the costs of education and unobserved village characteristics. We use the number of middle schools in the township and the number of high schools in the country to represent the costs of education.[5] More schools in a specific region imply a higher density of schools, which reduces the chances for students to stay in the school overnight, so families save money on children's boarding costs. As population growth slows down in China, school facilities have begun to show a sign of surplus and competition among schools has emerged to attract students. A direct result of the competition could be tuition and fee cuts. Therefore, we expect that both variables have positive effects on children's school attainment.

All our baseline regressions are estimated with the village fixed effects added. In rural China, schools are jointly supported by the central government (that is responsible for teachers' salaries) and local communities (that are responsible for other expenses) including villages (in the case of primary schools), towns and counties (in the case of middle schools and high schools). The village fixed effects control unobserved village and township and county characteristics (each county only has one village in our sample) as well as the potential group effects in individuals' school attainments.

3.2 Results based on the primary school cohort

We use the primary school cohort to study the impacts of health shocks before and in primary school on the decisions to finish primary school and to enter middle school, respectively. For the first decision, the entire sample of the cohort will be used; but for the second decision, a sub-sample of people who finished primary school will be used. In addition, we have deleted the observations in villages where all the people in the primary school cohort finished primary school when we study the decision to finish primary school. We do the same when we study the decision to enter middle school, that is, we have deleted the observations in villages where all the people in the sub-sample entered middle school. There existed strong village-based group effects in our sample. The average within-village standard deviations of the probabilities to finish primary school, enter middle school, finish middle school, and enter high school were 0.24, 0.34, 0.41, and 0.37, respectively, but the between-village standard deviations of the village averages of the four probabilities were 0.87, 0.77, 0.48, and 0.20, respectively. That is, between-village variations were larger than within-village variations for all the cases except the decision to enter high school. It is far from enough to control the large between-village variations only by the numbers of middle

[5] The number of primary schools did not have any variations over time in all but one village so we ignore it.

schools and high schools. The village fixed effects are critical for us to get consistent estimates for the effects of health shocks as well as for those of the other variables. Under this circumstance, keeping intact the observations in villages with no variations in school attainments will result in downward biases in our estimates for the impact of health shocks because the village fixed effects eliminate the differences between shock-hit and shock-free families. In reality, if a village could guarantee that all its children to finish a specific stage of education throughout the whole period of 1987-2002, this village must have out strong emphasis on education and deserves special treatment in our regressions.[6]

We put both the number of middle schools and the number of high schools in the regressions. These two variables take values in the year when a person was supposed to enter middle school, not the year when he was supposed to enter primary school, in order to obtain observations for every person (remember there are people who entered primary school before 1987 in the primary school cohort so we do not know the situation when they entered primary school). For the decision to enter middle school, the number of middle schools controls for the costs of middle school education, and the number of high schools controls for the expected costs of high school. For the decision to finish primary school, both variables control for the expected costs of future education. To the extent that more education is always an option, expected costs of future education matter for one's decision to finish his current stage of education.

[Table 2 about here]

The results of the decision to finish primary school and to enter middle school are reported in R1 and R3 of Table 2. In R1, 86.5% of the sample finished primary school, and the share of people with a health shock before and in primary school was 4.3% and 8.2%, respectively. In R3, 88.1% of the sample entered middle school, and the share of people with a health shock before and in primary school was 4.8% and 8.2%, respectively. Both regressions show that a health shock in primary school has a significantly negative impact on school attainment, but a health shock before primary school does not.[7] By the estimated coefficients, a person would lose 12.3% of his

---

[6] Deleting observations from such villages is but one alternative. In our household fixed-effect estimations, we will adopt a mixed panel strategy to deal with this issue. There were also occasionally one or two villages where all the children did not finish a certain stage of education. We do not delete observations in these villages.

[7] This result confirms our earlier doubt on the descriptive result that a shock before primary school raises a person's chances to finish primary school. This positive effect is indeed caused by the lack of control on personal,

chances to finish primary school and 8.1% of his chances to enter middle school if his family had an adult member being seriously sick in his primary school age. The insignificant results of shocks before primary school may be caused by the small percentage of people with such shocks. However, they may also reflect a real phenomenon. A health shock tightens up a family's liquidity constraint. Stopping children's education may be at the bottom of its list of means to release the constraint because investment in education pays off in the future. Since literacy is a necessity for even an agricultural laborer in rural China, a health shock before one's primary school age will not stop his family to send him to school.[8]

The primary school cohort has people whose primary school education age is not fully covered by our data. This has a drawback that we have miscoded people with shocks as shock-free observations so our estimates for the effect of shocks happening in primary school age are underestimated. To make a correction, we restrict the sample for both decisions to only include people of 13 to 21 years old in 2002 so we observe the entire primary school period of every person.[9] The estimation results are presented in R2 and R4 of Table 2 for the decision to finish primary school and to enter middle school, respectively. In the sample that R2 is based, 82.8% finished primary school, and 4.6% and 9.5% experienced a shock before and in primary school, respectively. In the sample of R3, 86.8% entered middle school, and 6.4% and 11.4% experienced a health shock before and in primary school, respectively. So we indeed capture more shocks. The effects of shocks before primary school are still insignificant. For shocks in primary school, their effect has indeed increased for the decision to enter middle school, but it has turned insignificant for the decision to finish primary school. However, this insignificant result may be caused by the small sample size of R2. The number of observations of R2 is 476, which is barely more than half of that of R1.

Combining the estimates of R1 and R3, one can imply that the total effect of a health shock in one's primary school age is to reduce his chances to enter middle

---

family, and village characteristics.

[8] On the other side of the story, a health shock in one's primary school age can have a significantly negative effect because literacy does not necessarily require a person to finish primary school.

[9] This new sample still does not allow us to observe everyone's entire pre-school period so the estimates for pre-school shocks (and to a possibly lesser extent, the estimates for primary school shocks) are biased. However, it is difficult to make a correction on this problem because we will have to reduce our sample to people aged 13-15 in 2002, which is too small to get reliable estimates.

school by 19.4%.[10] This is a very large effect and far larger than the gap between the dropout rates of the shock-hit and shock-free groups in the primary cohort (Table 1). Indeed, it is equivalent to 80.5% of the dropout rate of the shock-hit group.

This result can be compared with the total effect that we have directly estimated in R5 of Table 2. This regression uses the whole sample of the primary school cohort and studies the unconditional decision to enter middle school. While the coefficient of the health shock before primary school is still insignificant, the coefficient of the health shock in primary school is highly significant and indicates that a person's chances to enter middle school decreases by 18.3% because of the shock. This estimate is only 1.1 percent points away from our estimate inferred from the results of the sequential response model. Although the sequential response model is still the preferred model to describe educational decisions, this shows that the biases of the unconditional model are not serious.

Among the control variables, father's education is shown by all the five regressions to have a significantly positive impact on a child's school attainment, but the magnitude of the impact is relatively small: the largest is the total effect under R5, which is 3.3%. There is a significant gender gap in all the regressions but R2.[11] The size of the gap is large for the conditional and total effect for entering middle school, which are 5.8% and 7.6%, respectively. All the other control variables are insignificant except the number of older sisters and the number of high schools that are significant in R4 and mother's education and father's occupation in R5.[12]

3.3 Results based on the middle school cohort

Turning to the middle school cohort, we want to use it to study the decisions of finishing middle school and entering high school. The first decision uses the whole sample of the middle school cohort, and the second decision uses a sub-sample of it with people who had finished middle school. We are not only interested in health shocks happening in the middle school period, but also those before and in the primary school period because we want to study the long-term effects of the shock among which the screening effect is an important component.

[Table 3 about here]

---

[10] This is calculated by 100% - (100% - 12.3%)*(100% - 8.1%).

[11] This may be another sign of the small sample size of R2.

[12] The sign for the number of older sisters contradicts to, but the signs for the other three variables are consistent with our earlier expectations.

The results of the baseline regressions are presented in R1 and R3 in Table 3. In the sample of R1, 66.6% finished middle school, and the percentages of people who experienced a shock before primary school, in primary school, and in middle school were 2.4%, 5.8%, and 7.7%, respectively. The four corresponding figures for R3 are 39.9%, 2.8%, 7.1%, and 8.3%. Primary school shocks are shown by both regressions to have a positive but statistically insignificant effect on school attainment. In contrast, middle school shocks are shown by both regressions to have a negative but also statistically insignificant effect. The results for shocks happening before primary school are more interesting: they increase a person's chances to finish middle school by 13.9%, but reduce his chances to enter high school by 23.2%. The first result provides evidence for the existence of the screening effect, but the second result shows that this effect only lasts to middle school graduation. If we take these results seriously, then we can reasonably to believe that families must have endured a long period of hardship to support their children's education since they received a shock when the children were in their pre-school age, so nine years of compulsory education became the ultimate goal for their children's education. However, the small number of shocks before primary school in both regressions casts doubts on the stability of these results.

The middle school cohort does not cover the entire middle school age of people who were older than 26 in 2002. It also misses part of or the entire primary school age of people who were older than 21 in the same year. To check the robustness of our results for the shocks in primary and middle school, we thus rerun R1 and R3 with people between 16 and 21 years old (inclusive). The results are presented in R2 and R4 of Table 3. In the sample of R2, 65.5% finished middle school, and the percentages of people with shocks before primary school, in primary school, and in middle school were 4.7%, 9.1%, and 8.6%, respectively. In the sample of R4, 48.9% entered high school, and the corresponding percentages for the incidence of the shocks were 5.2%, 11.5%, and 9.3%, respectively. The results, however, qualitatively replicate those of R1 and R3, and the estimates for shocks before primary school become even larger. This last result, though, gives us confidence on the existence of the screening effect in the case of shocks before primary school because the percentage of such shocks is relatively large in both regressions.

There are two reasons for primary school shocks not to have an effect on middle school education. The first is that those shocks have already functioned once by preventing a substantially large number of people from entering middle school. The second is related to the mixture of the screening effect and the long-term negative impact of the shocks. Under the screening effect, kids who survive a shock and enter middle school must have above-average intelligence or their families much have high expectations on them, so they are less likely to fail in middle school. On the other hand, experiencing a shock reduce a family's capability to afford its children to complete middle school. For some kids, the screening effect dominates, but for others, the negative effect dominates, so on average shocks in primary school do not show significant impact on middle school education.

There are also two reasons for middle school shocks not to have an effect. The first is that primary school education itself can be a screening device for a family to find out whether its children have the academic potential for more education. In most places, a student needs to take a comprehensive test before he graduates from primary school. The score of this test can be used as a good indicator for a student's academic potential. The fact that a child already enters middle school shows that the family believes that he has that potential, inferring either from his test score or from his academic performance in primary school, or from both. As a result, people who have already entered middle school have higher resistance to the adverse effects of health shocks. The second reason is related to the relative costs of middle school education. The Chinese law allows people to work only after they turn 16 years old. The value of middle school education thus is high, especially if people want to find non-agricultural jobs. By the first reason, a child who is able to enter middle school has relatively high academic potential, so the risk of a failure is relatively low. Combining these two factors leads to the conclusion that investing in middle school education has low relative costs. This can be contrasted with investment in primary school education. For a child in primary school, there is an uncertainty that whether he could eventually finish primary school because there is no definitive test for his academic potential. Even when a family wants to invest in the child, there is a long way ahead along which unexpected shocks would arrive and force the family to stop his education. So the relative costs of investing in primary school education are higher.

4. Robustness tests

In this section, we conduct four robustness tests on our baseline results. Our analysis in the last section shows that health shocks' impacts on finishing primary school and entering middle school are the most robust results. So we will only conduct our robustness tests on these results.

4.1 The 1987-1996 sub-sample

Our first test is to correct the problem of the weak power of using income of 1987 to predict school attainment in 2002. We estimate a smaller sample of people who would finish primary school between 1988 and 1995. To still allow one year for a shock to take effect and for people to drop back in, this means that we are studying the sub-sample of the period of 1987-1996. The year 1996 is chosen for two reasons. One is related to the reports of major health shocks. As Figure 1 showed, there was a sharp increase of the number of reported major illnesses since 1996 while the number of reports before 1996 was more uniform across years. The other reason is related to improvements of healthcare services and facilities happening in the sample villages in the mid-1990s. As we will show soon in Figure 4, there were significant improvements starting in 1996. Better healthcare services and facilities reduce the relative costs of treating a sick family member, so the intensity of illness reports might change. Using the sub-sample of the early years also brings us a benefit to reduce the problem caused by behavioral changes in families' education investment. Our data covers a long period of time during which China experienced fast economic growth and policy changes, so it is natural to expect that families would change their pattern of education investment.

[Table 4 about here]

We perform three regressions using the 1987-1996 sample, one for the decision to finish primary school, one for the conditional decision to enter middle school, and the last for the unconditional decision to enter middle school. Their results are presented in Table 4. Instead of dropping the observations in villages with all people finishing the stage of education being studied, we keep all the people in the sample and group these villages together as the reference group for the other village dummies in order to increase the number of the observations. There are some differences between the current results and those based on the whole sample. A shock before primary school is shown to have a significant and positive impact on the probability to finish primary school and the unconditional probability to enter middle school. However, these estimates may not be credible because only 1.3% of the people in the 1987-1996 sample experienced a shock before they entered primary school.[13] The estimates for a

[13] This small percentage is a result of the truncation of the sample period. Many people's preschool period was before 1987 so their families' health history is unknown. We nevertheless keep the variable in the regressions

shock in primary school provide weaker results than those obtained in the whole sample. While they are still significant and negative for the decision to finish primary school and the unconditional decision to enter middles school, their magnitudes are smaller than in the whole sample. In addition, a shock does not have any significant effect on the conditional decision to enter middle school. These weaker results may be caused by the higher percentage of miscoded shocks in the 1987-1996 sample. We do not know the family health history before 1987 so people who entered primary school before 1987 are all coded as having not experienced any health shock. While the number of such miscoded observations is the same in both samples, their percentage is larger in the 1987-1996 sample than in the whole sample.

The results for the 1987 per-capita income have been slightly improved. While it is not significant in any regression using the whole sample, the 1987 income is now shown to have a significantly positive effect on the decision to finish primary school. Two other results of the control variables that are worth discussion are the estimates for the number of middle schools and the number of high schools. Both are significantly positive in the regression for the unconditional decision to enter middle school and their magnitudes are relatively large. These are the strongest results that we have obtained regarding the cost of education. It seems that the cost of education was more likely to be an impediment to school attainment in the earlier than in more recent years.[14]

4.2 2SLS estimations

Our second test is to instrument health shocks by a set of community variables that account for the healthcare conditions in the community that a family resides. One possible explanation, other than income growth, for the pattern shown in Figure 1 is that healthcare facilities and services began to improve in the mid-1990s. Improvements to these facilities lowered the relative costs of treatment so families were induced to be more likely to treat their sick members. On the other hand, health facilities and services do not directly affect a family's educational decisions other than through affecting the health of its members. So using healthcare facilities and services to instrument the reported shocks is a sensible approach. Figure 4 shows the trend of three variables representing, respectively, the average number of hospitals in the

---

because shocks before and in primary school may be correlated with each other.

[14] The estimate for number of high school is significantly negative for the decision to finish primary school, but its magnitude is only 0.3%.

township, the average number of hospitals in the county, and the number of villages with a healthcare plan in each year.[15] While there was no increase in the number of hospitals in the township, there was a significant jump of the number of hospitals in the county from 1996 to 1998. There was also a significant increase in the number of villages with a healthcare plan in the 1990s, but the sharpest increase still happened in 1996 to 1998. These trends generally match the trend of illness reports shown in Figure 1. Consequently, we use the number of hospitals in the township, the number of hospitals in the county, and a dummy variable indicating whether a village had a healthcare plan as the instruments for health shocks. To obtain data for all the people, we measure these three variables in the year when a person was supposed to finish primary school.

[Figure 4 about here]

The second-stage results of the 2SLS estimations are presented in Table 5. To save space, results of the control variables are not shown. We now go back to the 1987-2002 sample of the primary school cohort and still exclude the people in villages where all people passed the stage of education under study. For easy tractability, the linear probability model instead of the probit model is estimated in both stages of estimation. In the first stage estimation, household fixed effects instead of village effects are estimated (so time-invariant family variables are dropped). The R^2's of the first-stage estimation are high. For example, in the regression for the decision to finish primary school, the R^2 of the first-stage estimation for shocks before primary school is 0.85, and that for shocks in primary school is 0.79. As for the second-stage estimations, the results for the two shocks are mostly consistent with the baseline results. Shocks before primary school do not have a significant impact, but shocks in primary school significantly reduce a person's probability to finish primary school and his (unconditional) probability to enter middle school. The effect on the decision to finish primary school has the same magnitude as that of the baseline estimate, but the

[15] There are several kinds of healthcare plans in rural China. The most fundamental one is the village-based rural cooperative healthcare system. It was almost universal during the commune period, but dismantled in most villages since the rural reform took place as its finance depended on the commune system. In some advanced regions, the village-based system has been replaced by the township-based system that provides limited reimbursements to clinic visits. In recent years, the central government has begun to promote a new sort of cooperative healthcare plan that is based on voluntary participation and pools at the county level. On top of those plans, some villages have also joined limited commercial healthcare plans. In our study, we code a village as having a healthcare plan as long as it had any of the above plans regardless of the benefits they provided.

total effect on the decision to enter middle school is smaller than the baseline estimate. Lastly, the conditional effect on the decision to enter middle school is similar to that obtained in the 1987-1996 sample, which is insignificant.

[Table 5 about here]

4.3 Panel estimations with household fixed effects

Our third test is to do household fixed-effect estimation on the sub-sample of the primary school cohort with people who had siblings in the same cohort. The household fixed-effect estimation can bring us two benefits, both reducing the potential correlation between the incidence of health shocks and the error term. One is that it effectively controls family preferences for children's education as well as the persistence component of household income. The other is related to illness reporting. It is not only a matter of income, but also a matter of family preferences whether to treat a sick family member. The household fixed effects can provide effective control on the persistence component of family preferences.

[Table 6 about here]

Table 6 is comprised of three panels, each presenting the results of a set of fixed-effect regressions. Time-invariant family variables are dropped in each regression, and results of the control variables are not shown to save space. Panel A reports the main results of three regressions using the sub-sample of the primary school cohort that drops the observations in villages with all people passed the stage of education under study. However, none of them provides significant results for the health shocks. This may have something to do with behavioral changes in educational investment in the sample period. There were more reports of shocks in the later years, but families also tended to have higher awareness of the importance of children's education, so they might not reduce their children's education even if they had a health shock. This leads us to study the 1987-1996 sub-sample again. The results are presented in Panel B of Table 6. We have kept all the observations but dropped the variable indicating the shock before primary school because there are too few observations with this shock. Now the estimates of a shock in primary school are significant and negative for the decision to finish primary school and the unconditional decision to enter middle school. Both are quite large, being 30.1% and 25.9%, respectively. Comparing these two results with previous results, especially those in Table 4 that used the same sub-sample, one realizes that these two estimates are too large. One possible reason is that we have included observations in villages that have done very well in providing education so all their children finished the respective stage of education under study. In these villages, there were surely no variations of educational achievements among children from the same family no matter whether they experienced a health shock or not. In other words, families in these villages may have a different pattern of behavior in educational investment than families in other villages. To deal with this issue, we adopt a mixed panel method that only estimates the fixed effects of families in villages that did not have 100% primary school graduation rate. In the framework of the dummy-variable approach to fixed-effect estimation, this is equivalent to only adding family dummies to these families and treating families in villages with 100% primary school graduation as the reference group. The results are presented in Panel C of Table 6. The estimates for the shock in primary school are now much closer to those reported in Table 4. In addition, each of the three R^2's is larger than its counterpart in Panel B, so these estimates are more reliable than those in Panel B.

There are two questions about the estimates in Panel C, though. One is why they are considerably smaller than their counterparts in Panel B, and the other is why the total effect on entering middle school is smaller than the effect on finishing primary school. The answer to the first question is related to the correlation between the incidence of health shocks and the chances of a family being located in a village with a 100% primary school graduation rate. This correlation seems to be negative. While the incidence of shocks was 7.4% in villages with lower rates of primary school graduation, it was 5.7% in villages with 100% primary school graduation.[16] The fixed-effect panel estimation takes the families with changes in their status of having shocks as the treated group, and those without changes as the control group. So there is a larger percentage of families in villages with 100% primary school graduation that are treated as in the control group than in villages with smaller rates of primary school graduation. As a result, the effect of the shock is inflated. The mixed panel estimations in Panel C correct this problem by treating families in villages of 100% primary school graduation as a homogenous group and comparing all the other families with it.

The answer to the second question lies in the fact that a shock happening one's primary school age has a weak screening effect, that is, it increases his chances to enter middle school given that he finishes primary school. In fact, if we calculate the

[16] The correlation coefficient between the shock dummy and the dummy for 100% primary school graduation is -0.032, but is not statistically significant.

total effect by using the estimates for the decision to finish primary school and the conditional decision to enter middle school, we come up with a number of -12.7%, which is only 0.1 percentage points away from the estimate provided in Panel C of Table 6.

4.4 Dealing with the sibling effect

Our baseline results showed that having siblings does not have systematic and significant impacts on a person's chances to fare through a family health shock. However, the numbers of different siblings may not be enough to control the sibling effect that we described in the introduction. Specifically, our results may have heavily depended on the comparison between the older and younger children within the same family. This is no more evident in our household fixed-effect estimation. This estimation relies on families that had older children without experiencing a shock but younger children experiencing a shock to identify the effects of the shock,[17] so it is unable to distinguish whether the found effects are from health shocks or from the sibling effect. To deal with this issue, we rerun our baseline regressions using an alternative sample that includes all the children from the shock-free families and the first child experiencing a shock in the shock-hit families, but excludes all the other children in the latter kind of families. We still study the decisions to finish primary school and to enter middle school. The main results are presented in Table 7.

[Table 7 about here]

The first column of the table presents the main results of the probit model for the decision to finish primary school. Neither of the two marginal effects is significant. However, the point estimate for shocks in primary school is 0.55 and statistically significant at the 10% significance level (the standard error is 0.29). So the insignificant marginal effect has something to do with the method that marginal effects for dummy variables are calculated. To check our results, we rerun the regression using the linear probability model. In the results presented in the second column of Table 7, shocks in primary school are shown to have a statistically significant effect on a person's chances to finish middle school although it is smaller than what we found in the baseline regressions.

---

[17] Since health shocks are permanent to a family, younger children are automatically coded in our study as experiencing a shock if older children in the same family experienced one. So the only case in which children of the same family are coded differently is when the older children did not experience a shock, but the younger ones did.

Turning to the decision to enter middle school, the conditional effect of health shocks is still insignificant (column 3 of Table 7), but the total effect is (column 4 of Table 7), which are consistent with the results of our other robustness tests. In addition, the magnitude of the total effect is close to what we have found in our other robustness tests.

Judging by the above results, we conclude that the sibling effect does exist. However, it is premature to conclude that our other results exaggerate the effects of health shocks. The comparison between older and younger children within the same family may have picked up the sibling effect, but excluding such comparison may also overlook the real effects of health shocks. After all, there is a possibility that families would prefer to stop older children's education facing an income shock (Jacoby and Skoufias, 1996). Therefore, it is the best to believe that the effects of health shocks are between what we have obtained in the current test and what we have obtained in our baseline regressions. It is worth emphasizing that the lower bound of the effects is still substantial. Our direct estimate shows that the total effect of health shocks is to reduce a person's chances to enter middle school by 11.3%, and our calculation using the conditional estimates provides a figure of 10.6%.

5. Conclusions

Using a long panel of household and individual data, this paper has studied the impacts of adults' major illnesses on children's school attainment in rural China. Our baseline results show that children are the most vulnerable if an adult in their family has a major illness in their primary school age. The total effect of such a shock is to reduce a child's chances to enter middle school by 19.4%. The estimates obtained by our sensitivity tests are smaller, mostly ranging between 11% to 15%, but are still statistically significant and economically meaningful. We have also found weak evidence for the existence of the screening effect of health shocks. Related to this finding is that our results suggest that the negative impacts of health shocks are related to the temporary liquidity constraints caused by the shocks because shocks happening in one stage of a person's education do not have an effect on his next stage of education. The important thing, however, is that temporary liquidity constraints can lead to permanent school dropouts because it is difficult for a child to drop back in after being out of school for one or more years.

Although China set a law for compulsory nine year education in as early as 1986, the record shown in our sample has not been impressive. Among people who entered primary school in or after 1986, only 58.4% finished middle school (i.e., completed 9 year education). Our results have an important policy implication for improving education in rural China. Most government interventions are currently concentrated in providing better school facilities and more qualified teachers. While this is important and partly confirmed by our results, more attention should be paid to farm households' weak abilities to deal with unexpected risks among which health shocks are the most important. Providing farm households insurance against major illnesses will especially benefit children of primary school age because they are the most vulnerable group. This can be combined with a policy of providing educational loans to shock-hit households to release their temporary liquidity constraints. Because dropout in primary school will most likely to result in permanent deficiency in a person's educational attainment, providing health insurance and shock-related educational loans will bring large benefits to the society.

One potential improvement to our paper is to directly study the role of credit constraints in causing the negative effects of health shocks on education. The hard task is to find an objective measure of credit constraints. Relying on self-reported answers can be misleading. It waits for better survey instruments to get better measurements.

References

Brown, Philip (2006). "Parental Education and Investment in Children's Human Capital in Rural China." *Economic Development and Cultural Change*, 54: 759-790.

Cameron, Stephen and James Heckman (1998). "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy*, 106: 262-333.

Gan, Li, Lixin Xu and Yang Yao (2006a). "Finance, Village Governance and Consumption Smoothing in Rural China." Manuscript, the World Bank.

Gan, Li, Lixin Xu and Yang Yao (2006b). "Health Shocks, Village Governance and Farmers' Long-term Income Capabilities." Manuscript, CCER, Peking University.

Jacoby, Hanan and Emmanuel Skoufias (1996). "Risk, Financial Markets, and Human Capital in a Developing Country." *The Review of Economic Studies,* 64: 311-335.

Gertler, Paul and Jonathan Gruber (2002). "Insuring Consumption against Illness." *American Economic Review*, 92: 51-70.

Li, Jing, Justin Lin, and Yang Yao (2002). "Credit Constraints, Land, and Children's Education in Underdeveloped Regions." *Chinese Journal of Population Science*, 2002(6): 10-26.

Rosenzweig, Mark and Kenneth Wolpin (1993). "Credit Market Constraints, Consumption Smoothing and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India." *Journal of Political Economy,* 101: 223-244.

Schultz, Paul (1999). "Health and Schooling Investments in Africa." *Journal of Economic Perspectives*, 13: 67-88.

Thomas，Duncan, et al. (2004). "Education in a Crisis." *Journal of Development Economics*, 74: 53-85.

Townsend, Robert (1994). "Risk and Insurance in Village India." *Econometrica*, 62: 539-592.

Table 1. Statistics of the primary school and middle school cohorts

Panel A: Primary school cohort (age 13 – 26 in 2002)

	Total	By shocks before primary school		By shocks in primary school	
		Shock-hit group [1]	Shock-free group [1]	Shock-hit group	Shock-free group
# observations	1213	58	1155	120	1093
# completing primary school	1077	55	1022	105	972
Primary school dropout rate [2]	11.2%	5.2%	11.5%	12.5%	11.1%
# entering middle school	966	47	919	91	875
Perc. entering middle school [3]	89.7%	85.5%	89.9%	86.7%	90.0%
Total dropout rate [4]	20.4%	19.0%	20.4%	24.2%	19.9%

Panel B: Middle school cohort (age 16 – 29 in 2002)

	Total	By shocks before primary school		By shocks in primary school		By shocks in Middle school	
		Shock-hit group	Shock-free group	Shock-hit group	Shock-free group	Shock-hit group	Shock-free group
# observations	1229	29	1200	73	1156	95	1134
# completing middle school	652	18	634	46	606	54	598
Middle school dropout rate [5]	46.9%	37.9%	47.2%	37.0%	47.6%	43.2%	47.3%
# entering high school	260	5	255	23	237	19	241
Perc. entering high school [6]	39.9%	27.8%	40.2%	50.0%	39.1%	35.1%	40.3%
Total dropout rate [7]	78.8%	82.8%	78.8%	68.5%	79.5%	78.8%	82.8%

Notes:

1. The shock-hit group includes people whose families experienced a health shock when they were in a specific age period, and the shock-free group includes people whose families did not. Definitions are the same in the table.
2. Percentage of people who did not finish primary school in the primary school cohort.
3. Percentage of people who attended middle school in the group of people who finished primary school.
4. Percentage of people who did not enter middle school in the primary school cohort.
5. Percentage of people who did not finish middle school in the middle school cohort.
6. Percentage of people who attended high school in the group of people who finished middle school.
7. Percentage of people who did not attend high school in the middle school cohort.

Table 2. Results for decisions to finish primary school and enter middle school

	Finishing primary school			Entering middle school	
	R1	**R2**	**R3**	**R4**	**R5**
Constant	-0.232***	-0.336**	-0.129	-0.319***	-0.731***
	(0.080)	(0.146)	(0.084)	(0.121)	(0.134)
Health shock before primary school	0.028	0.020	-0.047	-0.078	-0.032
	(0.031)	(0.058)	(0.046)	(0.062)	(0.062)
Health shock in primary school	-0.123**	-0.094	-0.081*	-0.107**	-0.183***
	(0.060)	(0.084)	(0.045)	(0.054)	(0.065)
Gender	0.038**	0.023	0.043***	0.058***	0.076***
	(0.017)	(0.029)	(0.016)	(0.021)	(0.025)
# Older brothers	-0.002	0.024	-0.007	0.002	-0.010
	(0.015)	(0.026)	(0.015)	(0.017)	(0.024)
# Older sisters	0.000	-0.013	-0.009	-0.030**	-0.011
	(0.011)	(0.021)	(0.010)	(0.013)	(0.017)
# Younger brothers	0.011	-0.010	0.024	0.014	0.032
	(0.017)	(0.030)	(0.018)	(0.020)	(0.027)
# Younger sisters	-0.008	-0.025	-0.002	-0.012	-0.002
	(0.016)	(0.030)	(0.016)	(0.019)	(0.024)
1987 per capita income (1,000 Yuan)	0.016	-0.003	0.036	0.040	0.028
	(0.033)	(0.065)	(0.028)	(0.030)	(0.046)
Father's schooling years	0.017***	0.028***	0.016***	0.015***	0.033***
	(0.004)	(0.008)	(0.004)	(0.005)	(0.007)
Mother's schooling years	0.004	0.008	0.007*	0.002	0.011*
	(0.004)	(0.008)	(0.004)	(0.005)	(0.006)
Father's occupation in 1987	0.033*	0.040	0.025	0.001	0.064*
	(0.020)	(0.034)	(0.022)	(0.031)	(0.034)
# Middle schools in township	0.021	0.014	0.019	-0.010	0.024
	(0.014)	(0.031)	(0.015)	(0.022)	(0.022)
# High schools in county	-0.017	0.021	0.014	0.043***	-0.003
	(0.026)	(0.045)	(0.013)	(0.009)	(0.012)
# Observations	925	476	910	577	1096
Pseudo-R^2	0.209	0.206	0.106	0.214	0.225

Notes: The probit model is estimated for each regression. The dependent variable is whether to finish a certain stage of education. R1 and R5 are based on the whole sample of the primary school cohort (age 13-26 in 2002), R2 is based on a sub-sample of the cohort with people of age 13-21 in

2002, R3 is based on a sub-sample of people in the primary school cohort who had finished primary school, and R4 is a sub-sample of this sub-sample with people of age 13-21. The dependent variable of R5 is the unconditional decision to enter middle school. Observations in villages with all the people passing the stage of education under study are dropped in each regression. The coefficients are marginal effects evaluated at variables' sample means. Figures in the parentheses are the White robust estimates of the standard errors for the corresponding estimates. Results for the dummies indicating the year of school entrance and the village fixed effects are not shown. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.

Table 3. Results for decisions to finish middle school and enter high school

	Finishing middle school		Entering high school	
	R1	**R2**	**R3**	**R4**
Constant	0.105	0.441	-0.253	-0.655**
	(0.165)	(0.600)	(0.211)	(0.344)
Health shock before	0.139*	0.154*	-0.232**	-0.291**
primary school	(0.797)	(0.089)	(0.093)	(0.134)
Health shock in	0.064	0.054	0.045	0.157
primary school	(0.076)	(0.097)	(0.090)	(0.114)
Health shock in	-0.036	-0.043	-0.014	-0.071
middle school	(0.062)	(0.104)	(0.081)	(0.122)
Gender	0.053	0.001	-0.081	0.066
	(0.034)	(0.056)	(0.045)	(0.079)
# Older brothers	-0.071**	-0.122**	-0.027	0.056
	(0.036)	(0.037)	(0.052)	(0.102)
# Older sisters	-0.019	-0.046	0.066*	0.116*
	(0.024)	(0.035)	(0.034)	(0.062)
# Younger brothers	0.035	0.070	-0.028	-0.009
	(0.040)	(0.063)	(0.052)	(0.086)
# Younger sisters	-0.063**	-0.113**	0.040	0.016
	(0.032)	(0.057)	(0.041)	(0.074)
1987 per capita	0.082	0.136	0.145*	0.130
income (1,000 Yuan)	(0.082)	(0.133)	(0.085)	(0.168)
Father's schooling	0.017**	0.018	0.017*	0.014
years	(0.009)	(0.013)	(0.010)	(0.017)
Mother's schooling	0.012	0.023*	0.021**	0.042**
years	(0.008)	(0.013)	(0.010)	(0.018)
Father's occupation	0.049	-0.016	0.021	0.090
in 1987	(0.051)	(0.080)	(0.067)	(0.112)
# Middle schools in	-0.028	-0.006		
township	(0.023)	(0.115)		
# High schools in	0.008	-0.236	-0.004	0.076
county	(0.015)	(0.142)	(0.018)	(0.081)
# Observations	977	385	652	270
Pseudo-R^2	0.290	0.254	0.221	0.272

Notes: The probit model is estimated for each regression. The dependent variable is whether to finish a certain stage of education. R1 is based on the whole sample of the middle school cohort (age 16-29 in 2002), R2 is based on a sub-sample of people aged 16-21 in 2002, R3 is based on a sub-sample of people who had finished middle school, and R4 is based on a sub-sample of this sub-sample with people aged 16-21 in 2002. Observations in villages with all the people passing the stage of education under study are dropped in each regression. The coefficients are marginal effects evaluated at variables' sample means. Figures in the parentheses are the White robust estimates of the standard errors for the corresponding estimates. Results for the dummies indicating the year of school entrance and the village fixed effects are not shown. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.

Table 4. Results based on the 1987-1996 sub-sample [1]

	Finishing primary school [2]	Entering middle school	
		Conditional effect [3]	Total effect [2]
Constant	0.054	-0.007	-0.048***
	(0.047)	(0.007)	(0.011)
Health shock before primary school	0.068***	-0.001	0.067***
	(0.012)	(0.005)	(0.023)
Health shock in primary school	-0.119*	-0.001	-0.155**
	(0.064)	(0.003)	(0.078)
Gender	0.012	0.002	0.044**
	(0.018)	(0.003)	(0.021)
# Older brothers	-0.024	0.000	-0.008
	(0.016)	(0.001)	(0.021)
# Older sisters	-0.002	0.001	0.016
	(0.011)	(0.001)	(0.013)
# Younger brothers	-0.009	0.002	0.034
	(0.022)	(0.003)	(0.025)
# Younger sisters	-0.016	0.001	0.017
	(0.016)	(0.001)	(0.020)
1987 per-capita income (1000 yuan)	0.053*	0.001	0.051
	(0.030)	(0.002)	(0.045)
Father's schooling years	0.098**	0.001	0.020***
	(0.042)	(0.001)	(0.006)
Mother's schooling years	0.002	0.000	0.008
	(0.003)	(0.000)	(0.006)
Father's occupation in 1987	0.051***	-0.001	0.043
	(0.018)	(0.002)	(0.027)
# Middle schools in township	0.008	0.002	0.063*
	(0.007)	(0.002)	(0.038)
# High schools in county	-0.003***	0.003	0.023**
	(0.001)	(0.004)	(0.010)
# Observations	693	616	693
Pseudo-R^2	0.174	0.159	0.255

Notes:

1. The probit model is estimated for each regression. The dependent variable is whether to finish a certain stage of education. Coefficients are marginal effects evaluated at sample variable

means. Figures in the parentheses are White robust estimates for the standard errors. Village dummies and school entrance dummies are included in each regression but their results are not shown. Villages with all children passing the stage of education being studied are used as the reference group for the village dummies. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.

2. The sample includes people who were between 19 and 26 years old (inclusive) in 2002.
3. The regression is based on a sub-sample of the sample of the first regression with people who had finished primary school.

Table 5. Results of the 2SLS estimations [1]

	Finishing primary school [2]	Entering middle school	
		Conditional effect [3]	Total effect [2]
Health shock before primary school	0.012	-0.018	-0.011
	(0.060)	(0.062)	(0.066)
Health shock in primary school	-0.123***	-0.061	-0.123***
	(0.048)	(0.048)	(0.048)
# Observations	925	895	1096
Adjusted R^2	0.210	0.133	0.224

Notes:

1. The linear probability model is estimated in each regression. The dependent variable is a 0-1 binary variable indicating whether a person has finished a certain stage of education. Observations in villages with all the children passing the stage of education being studied are dropped in each regression. Health shocks before and in primary school are instrumented by three variables: a dummy variable indicating whether a village had any healthcare plan, the number of hospitals in the township, and the number of hospitals in the county, all measured in the year when a person was at the age to enter middle school. Household fixed effects are used in the first stage of regression. The control variables are the same as in the baseline regressions, but their results are not shown. Also not shown are the results of the school entrance dummies and the village fixed effects. Figures in parentheses are heteroscedasticity-corrected standard errors. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.
2. The sample used in the regressions is the primary school cohort with observations in villages with all the children passing the stage of education under study being dropped.
3. The regression is based on a sub-sample of the primary school cohort with people who had finished primary school. Again, observations in villages with all the children passing the stage of education under study are dropped.

Table 6. Results of the household fixed-effect estimations [1]

	Finishing primary school [2]	Entering middle school	
		Conditional effect [3]	Total effect [2]
Panel A: 1987-2002 sample [4]			
Health shock before primary school	0.108	-0.051	-0.047
	(0.082)	(0.088)	(0.381)
Health shock in primary school	-0.041	-0.038	0.070
	(0.054)	(0.060)	(0.108)
# Observations	923	829	923
Adjusted R^2	0.416	0.237	0.418
Panel B: 1987-1996 sub-sample (I) [5]			
Health shock in primary school	-0.301***	0.070	-0.259***
	(0.082)	(0.108)	(0.110)
# Observations	426	381	426
Adjusted R^2	0.587	0.406	0.537
Panel C: 1987-1996 sub-sample (II) [6]			
Health shock in primary school	-0.153***	0.027	-0.128*
	(0.053)	(0.070)	(0.076)
# Observations	426	381	426
Adjusted R^2	0.664	0.423	0.565

Notes:

1. The linear probability model is estimated in each regression with household and school entrance fixed effects. The dependent variable is a 0-1 binary variable indicating whether a person has finished a certain stage of education. Time-invariant family variables are dropped in each regression, and results of the control variables are not shown. Figures in parentheses are heteroscedasticity-corrected standard errors. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.
2. The sample used in the regressions is comprised of people in the primary school cohort who had siblings in the same cohort.
3. The regression is based on a sub-sample of the one defined in Note 2 with people who had

finished primary school.

4. Observations in villages with all the children passing the stage of education under study are dropped in each regression.
5. Household fixed effects are estimated for each household.
6. Household fixed effects are not estimated for households in villages where all children finished primary school.

Table 7. Estimations controlling the sibling effect [1]

	Finishing primary school [2]		Entering middle school [3]	
	Probit model	**Linear probit model**	**Conditional effect**	**Total effect**
Health shock before	-0.034	-0.055	0.012	0.014
primary school	(0.058)	(0.077)	(0.009)	(0.047)
Health shock in	-0.073	-0.095*	-0.012	-0.113*
primary school	(0.053)	(0.058)	(0.018)	(0.060)
# Observations	831	831	803	978
Pseudo-R^2/Adjusted R^2	0.215	0.210	0.151	0.245

Notes:

1. The dependent variable is a 0-1 binary variable indicating whether a person has finished a certain stage of education. The sample includes all the children in shock-free families and the first child experiencing a shock in shock-hit families. Observations in villages with all the children passing the stage of education being studied are dropped in each regression. The control variables are the same as in the baseline regressions, but their results are not shown. Results for the school entrance dummies and the village fixed effects are also not shown. * indicates the 10% significance level, ** indicates the 5% significance level, and *** indicates the 1% significance level.
2. The sample is the same in both regressions and is drawn from the primary school cohort. In the probit regression, marginal effects evaluated at variable means are reported and their White-robust standard errors are presented in the parentheses. In the linear probit model, heteroscedasticity-corrected standard errors are reported.
3. The probit model is estimated in both regressions. The first regression is based on a sub-sample of the primary school cohort with people who had finished primary school, and the second regressions draws on the primary school cohort. Marginal effects evaluated at variable means are reported and their White-robust standard errors are presented in the parentheses.

Figure 1. Number of major illnesses in the sample

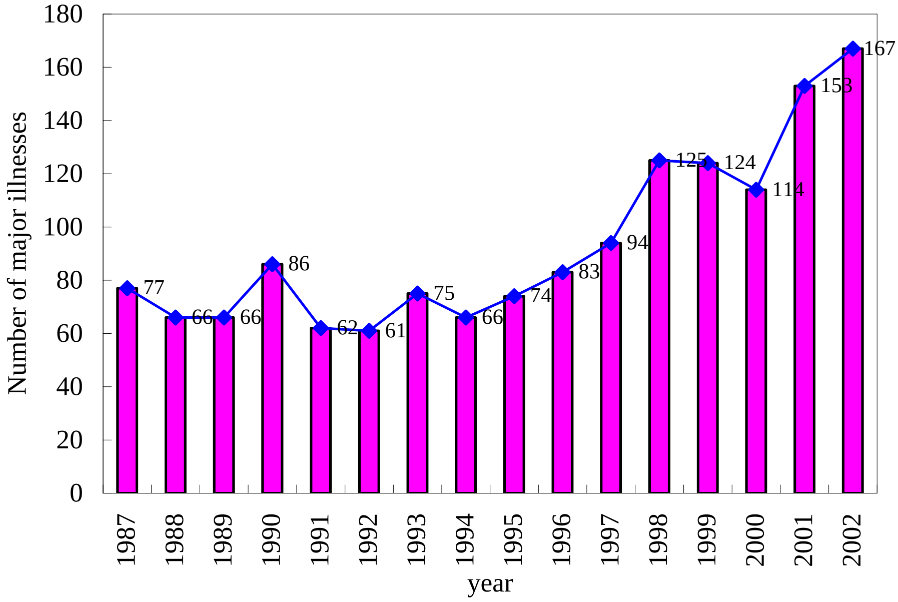
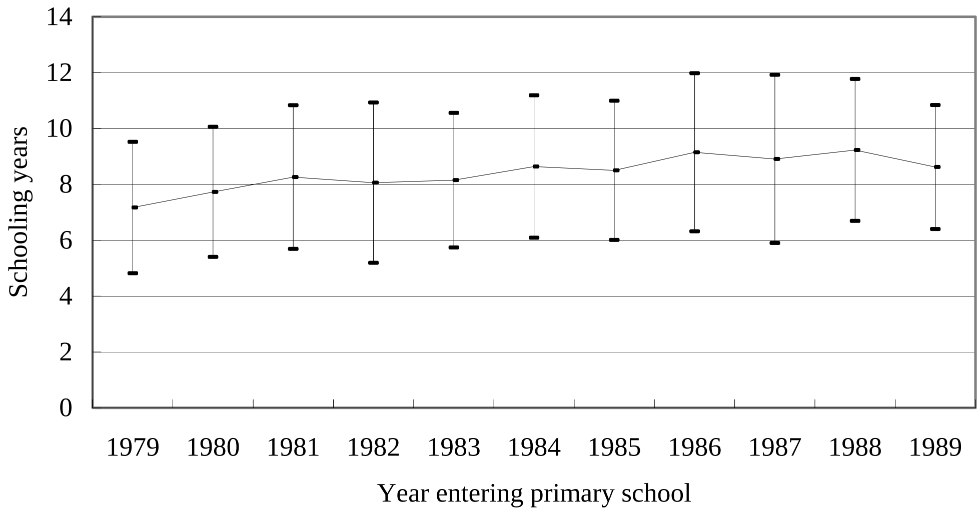
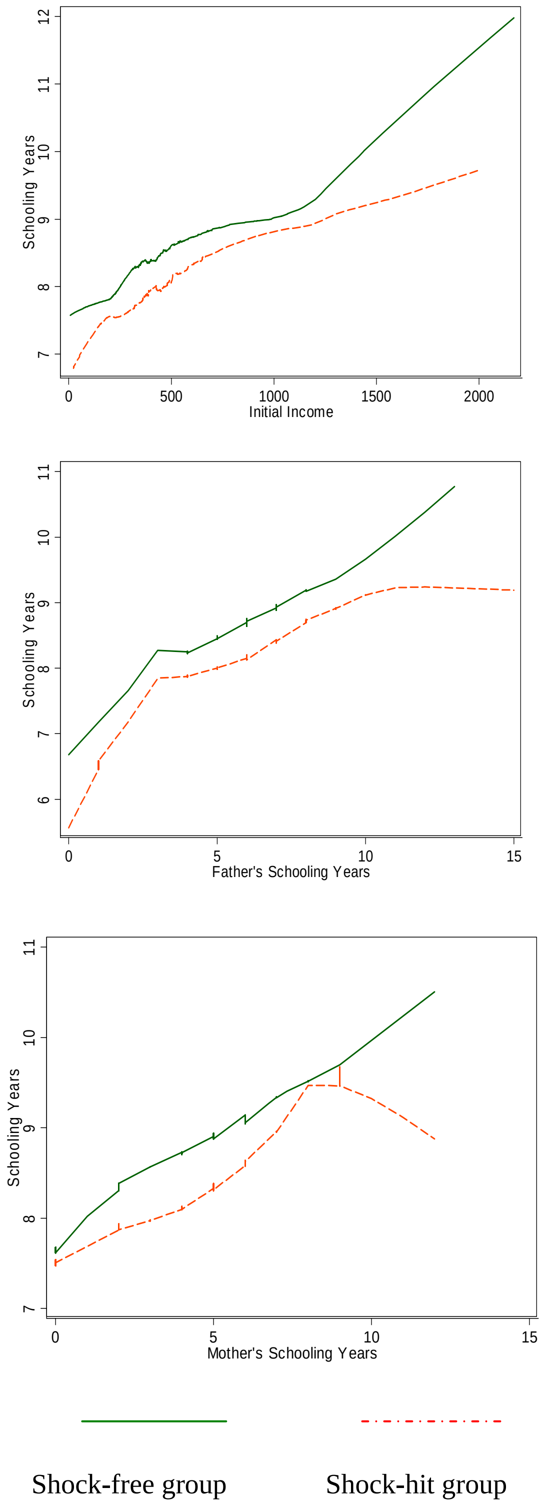


Figure 2. School attainment for people entering primary school in 1976-1989



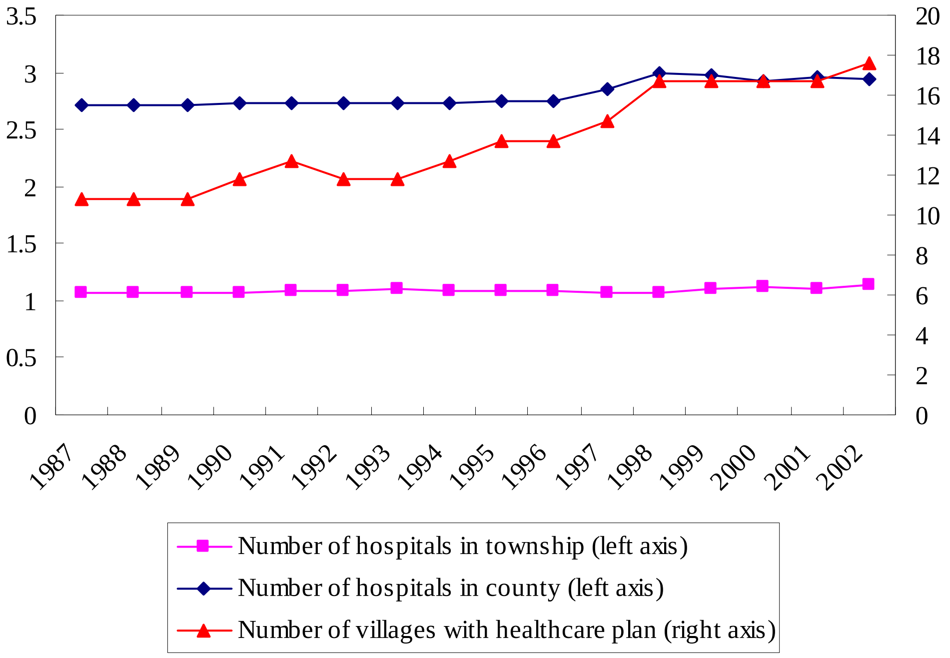
Notes: The figure is based on the sample of 1,275 persons between 19 and 32 years old in 2002 who were supposed to graduate from high school in the period of 1988-2001. The line indicates average educational attainment by 2002 of people entering primary school in a specific year, and each bar indicates the range of one standard deviation above and below the average for each year.

Figure 3. Impacts of health shocks on school attainments by initial income and parents' education



Notes: The chats are based on the sample of 1,275 persons between 19 and 32 in 2002 who were supposed to graduate from high school in the period of 1988-2001. The lines in the figures are moving-average trends of actual data.

Figure 4. Improvements of healthcare facilities and services



Notes: The number of hospitals in township and county, respectively, is the sample average in each year.