Early Health Shocks, Intrahousehold Resource Allocation, and Child Outcomes

Web Appendix Not for Publication*

Junjian Yi
National University of Singapore

James J. Heckman
University of Chicago

Junsen Zhang
Chinese University of Hong Kong

Gabriella Conti
University College London

December 1, 2014

---

*Junjian Yi, Department of Economics, National University of Singapore; email: `junjian.yi@gmail.com`. James Heckman: Department of Economics, University of Chicago; email: `jjh@uchicago.edu`. Junsen Zhang, Department of Economics, Chinese University of Hong Kong; email: `jszhang@cuhk.edu.hk`. Gabriella Conti, Department of Applied Health Research, University College London; email: `gabriella.conti@ucl.ac.uk`. The research was supported in part by the American Bar Foundation, the Pritzker Children's Initiative, the Buffett Early Childhood Fund, NIH grants NICHD R37HD065072 and R01HD54702, an anonymous funder, an European Research Council grant (DEVHEALTH 269874), a grant for the FIS from the Chinese University of Hong Kong, and a grant from the Institute for New Economic Thinking (INET) to the Human Capital and Economic Opportunity Global Working Group (HCEO)-an initiative of the Becker Friedman Institute for Research in Economics (BFI). The views expressed in this paper are those of the authors and not necessarily those of the funders or persons named here.

1 Appendix A: The Measure of Child Early Health Shocks

We discuss the measure of child early health shocks in this section. The CCTS questionnaire contains questions about the history of serious diseases for each child. The complete list of diseases include serious diarrhea, calcium deficiency, asthma, fracture, attention deficit disorder, heart disease, serious hearing difficulties, whooping cough, stammer, and serious eyesight problem. To minimize recall errors, this part of the questionnaire was answered by parents and children together. Furthermore, the parents and children were required to specify the timing and duration of each disease.[1] The most common disease reported is serious diarrhea. This fact is consistent with the case for children in developing countries (Strauss and Thomas, 1998). The average age when children suffered from serious diarrhea is 1.42. The duration of this disease is 1.54 months. For the case of both twin children suffered from the disease, 61.23% of the twins were suffered at the same time.

One assumption with the specification of our production function (Equation (1) in the paper) is that no contagious or spillover effect of child early health shocks exists in the production function. Does the serious diarrhea reported in our survey violate this assumption? Our answer is probably no.

There are two types of diarrhea. One is infectious and the other is non-infectious (Chen and Scrimshaw, 1983).[2] Viral gastroenteritis with rotavirus are the most common cause for infectious diarrhea. Fang et al. (2005) show that rotavirus diarrhea accounts for about 50% of the cases in children under five in China. In general, infections that cause diarrhea are highly contagious. However, the symptoms of diarrhea caused by viral gastroenteritis usually last just a few days (Chen and Scrimshaw, 1983). In the case of viral gastroenteritis, replacing lost fluid and salts is the only treatment needed.

---

[1] See Appendix C for the discussion on potential measurement error problems.

[2] We thank several discussions with Dr. Zhixiang Zuo in the Medical School at University of Chicago and with Dr. Linjun Wang in the Medical School at Michigan University.

The non-infectious diarrhea is usually caused by poisoning food, such as poisoning mushroom, and polluted water. The non-infectious diarrhea usually is very acute and needs medical treatments. If children suffering from non-infectious diarrhea could not get timely and appropriate medical treatment, the diarrhea becomes chronic, leading to ulcerative colitis, celiac disease, or irritable bowel syndrome. The chronic diarrhea can last much longer than the infectious one.

We believe that the type of diarrhea varying between twins is the non-infectious one caused by eating poisoning food or drinking polluted water.[3] On one hand, the duration of diarrhea reported in the survey is much longer than that in case of infectious diarrhea. On the other hand, the interviewers asked the parents about *serious* diseases during the survey. So the parents were unlikely to report the infectious diarrhea which lasted only for a short period and did not have medical treatment for their children.

We suggest a simple test to check potential contagious effects of early health shocks in the production function. If the early health shock is contagious, both twin siblings who shared a bedroom during childhood would more likely experience early health shocks than those who did not share a bedroom. The reason is that the rotavirus are more likely to spread to each other if they sleep in a same bedroom. We test this hypothesis by conducting a multinomial logit estimation. Specifically, we regress the number of twin siblings who suffered from early health shock on a dummy variable that indicates whether the twin children shared a bedroom. The results are reported in Table A1. We find that sharing a bedroom does not significantly affect the possibility that both twin children suffer from early health shock (columns (2) and (4)). Our evidence is inconsistent with the hypothesis of contagious effects. In addition, Table A1 shows that no covariate significantly predicts that only one child suffers from early health shocks (columns (1) and (3)). This result confirms our assumption of the randomness of the within-twin variation in early health shock.

---

[3] Chen, Ebenstein, Greenstone, and Li (2013) and Ebenstein (2012) show water pollution is an increasingly important threat to health in China.

Table A1: Multinorminal Logit Estimates of Sharing a Room and the Number of Children Suffered from Early Health Shocks

	Dependent variables: # Children Suffered from Early Health Shocks:			
	Only one	Both	Only one	Both
	(1)	(2)	(3)	(4)
Sharing a room	-0.006	0.008	-0.010	0.004
	[0.011]	[0.016]	[0.012]	[0.017]
Child birth weight (mean)			-0.013	-0.023
			[0.011]	[0.015]
Child age			0.001	-0.001
			[0.002]	[0.003]
Male twins			0.008	0.018
			[0.012]	[0.017]
Female twins			-0.006	-0.018
			[0.013]	[0.019]
Twin born at the first parity			0.025	-0.002
			[0.016]	[0.018]
Maternal age			-0.002	-0.001
			[0.001]	[0.002]
Maternal ethnicity (Han=1)			-0.002	-0.013
			[0.014]	[0.018]
Maternal school years			0.002	0.005**
			[0.002]	[0.003]
Maternal working sector (public=1)			-0.025	0.013
			[0.021]	[0.022]
Rural			-0.014	-0.010
			[0.011]	[0.015]
Household asset			-0.003	-0.005
			[0.003]	[0.005]
# Households	1,456	1,456	1,456	1,456

Note: The omitted category is that no one in a pair of twins suffered early health shocks. Columns (1) to (4) reports the marginal effects of the multinorminal logit estimates. Standard errors are in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%.

2 Appendix B: An Example of the Sensitivity of Parental Responses to Child Early Health Shocks Depends on Specific Functional Forms

2.1 An Example

We characterize the dependence of the results in the previous literature on choices of functional forms. To derive analytic results, we first specify the parental utility function as

$$U = c^{a_1} l^{a_2} [(q_i^{\rho} + q_j^{\rho})^{\frac{1}{\rho}}]^{1-a_1-a_2}, \tag{1}$$

where $0 < a_1, a_2 < 1$, and $\rho \leq 1$. The parameter ρ measures the degree of parental inequality aversion across twin siblings. For example, if $\rho = 1$, parents exhibit zero inequality aversion. When $\rho = -\infty$, parents exhibit infinite inequality aversion. The latter is called a Rawlsian case. The child quality is a combination of health and cognitive skills such that

$$q_{\iota} = (\theta_{\iota}^{H})^{\alpha_H} (\theta_{\iota}^{C})^{\alpha_C},$$

where $0 < \alpha_H, \alpha_C < 1$. We specify the production functions as

$$\begin{aligned} \theta_{\iota}^{H} &= (\eta_{\omega} \omega_{\iota}^{C})^{\gamma} (\beta_{\omega} \omega_{\iota}^{H} + \beta_e e_{\iota}^{H} + \beta_I I_{\iota}^{H})^{1-\gamma}, && (2) \\ \theta_{\iota}^{C} &= (\eta_{\omega} \omega_{\iota}^{H} + \eta_e e_{\iota}^{H})^{\gamma} (\beta_{\omega} \omega_{\iota}^{C} + \beta_I I_{\iota}^{C})^{1-\gamma}, && (3) \end{aligned}$$

where $0 < \gamma, \eta_{\omega}, \beta_{\omega}, \beta_I < 1$, and $\beta_e, \eta_e < 0$.[4] Comparing with the general production function (Equation 1) in the paper, we omit parental and child's characteristics. The specific production function form is used only for analytic convenience. By nesting a linear function into a CD function, we can easily derive the analytical comparative static result. At the

---

[4] This is because that e is defined as a negative shock.

same time, the result is general enough to demonstrate the two major observations.

Solving the optimization problem, we derive the difference between the self and cross-sibling effects of early health shock on family investment for child i as follows: [5]

$$\frac{\partial I_i^{H*}}{\partial e_i^H} - \frac{\partial I_i^{H*}}{\partial e_j^H} = \frac{\alpha_H}{\beta_I}\left(\frac{\partial \pi_i}{\partial e_i^H} - \frac{\partial \pi_i}{\partial e_j^H}\right)W - \frac{\beta_e}{\beta_I}, \tag{4}$$

$$\frac{\partial I_i^{C*}}{\partial e_i^H} - \frac{\partial I_i^{C*}}{\partial e_j^H} = \frac{\alpha_C}{\beta_I}\left(\frac{\partial \pi_i}{\partial e_i^H} - \frac{\partial \pi_i}{\partial e_j^H}\right)W, \tag{5}$$

where

$$W = \beta_\omega \sum_k \sum_\iota \omega_\iota^k + \beta_I \sum_k \sum_\iota I_\iota^k + \beta_e \sum_\iota e_\iota^H, \tag{6}$$

and

$$\pi_i = \frac{q_i^\rho}{q_i^\rho + q_j^\rho}. \tag{7}$$

We note that the sign of $\partial\pi_i/\partial e_i^H - \partial\pi_i/\partial e_j^H$ is unambiguously determined by parental preference. Specifically, $sign[\partial\pi_i/\partial e_i^H] = sign[-\rho]$ and $sign[\partial\pi_i/\partial e_j^H] = sign[\rho]$. If parents are sufficiently inequality averse ($\rho < 0$), then $\partial\pi_i/\partial e_i^H - \partial\pi_i/\partial e_j^H$ is positive. Otherwise, it is negative.

We now investigate parental investment strategy or the intrahousehold resource allocation effect. Equation (4) shows that the reinforcing or compensatory health investment in response to early health shocks is determined by not only parental aversion to inequality but also production technology. The sign of the first term on the right hand side of the equation is determined by parental preference, whereas the second term is determined by production technology. This result is consistent with our first observation listed above. By combining Equations (4) and (5), we note that parents could reinforce the human capital investment in one dimension and compensate in another dimension. Assume $\rho > 0$, parents reinforce investment in cognitive skills, but parental investment strategy with respect to health is undetermined. This result is consistent with our second observation listed above. There-

---

[5] The mathematical derivation and detailed discussion are presented in the section below.

fore, we conclude that intrahousehold resource allocation effects could have different signs depending on different types of human capital investment. The role of the intrahousehold resource allocation in the overall level of inequality in the economy is undetermined.

The effects of child early health shocks on parental consumption and labor supply are as follows:

$$\frac{\partial T^*}{\partial e_i^H} = -\frac{a_2\beta_e}{w[1-\gamma(1-a_1-a_2)]},$$

and

$$\frac{\partial c^*}{\partial e_i^H} = \frac{a_1\beta_e}{[1-\gamma(1-a_1-a_2)]}.$$

As $\beta_e < 0$, early health shock increases parental labor supply but decreases parental consumption. Thus, we conclude that exclusively focusing on the effects within twin siblings understates the total negative effect at the household level.

2.2 Mathematical Derivation of the Comparative Static Results

Parents maximize

$$\begin{aligned} \underset{c,l,I_{i,1}^H,I_{i,1}^C,I_{j,1}^H,I_{j,1}^C}{Max} U &= U(c,l,Q) \\ &= c^{a_1} l^{a_2} [(q_i^\rho + q_j^\rho)^{\frac{1}{\rho}}]^{1-a_1-a_2}, \end{aligned}$$

subject to the time constraint

$$l + T = 1,$$

and the budget constraint

$$\sum_k \sum_\iota p_I I_\iota^k + c = Y + wT \quad (k = H, C; \iota = i, j).$$

The child quality function is

$$q_\iota = (\theta_\iota^H)^{\alpha_H} (\theta_\iota^C)^{\alpha_C}.$$

The human capital production functions are

$$\begin{aligned}\theta_\iota^H &= (\eta_\omega \omega_\iota^C)^\gamma (\beta_\omega \omega_\iota^H + \beta_e e_\iota^H + \beta_I I_\iota^H)^{1-\gamma},\\ \theta_\iota^C &= (\eta_\omega \omega_\iota^H + \eta_e e_\iota^H)^\gamma (\beta_\omega \omega_\iota^C + \beta_I I_\iota^C)^{1-\gamma}.\end{aligned}$$

The Lagrangian is

$$\mathcal{L} = U(c, l, Q) + \lambda(Y + w - p_I \sum_k \sum_\iota I_\iota^k - c - wl).$$

The first-order conditions are

$$\begin{aligned}\frac{a_1 U}{c} &= \lambda && (8)\\ \frac{a_2 U}{l} &= \lambda w && (9)\\ (1-a_1-a_2)U\pi_i\alpha_H(1-\gamma)\beta_I\frac{1}{\beta_\omega\omega_i^H + \beta_e e_i^H + \beta_I I_i^H} &= \lambda p_I && (10)\\ (1-a_1-a_2)U\pi_i\alpha_C(1-\gamma)\beta_I\frac{1}{\beta_\omega\omega_i^C + \beta_I I_i^C} &= \lambda p_I && (11)\\ (1-a_1-a_2)U\pi_j\alpha_H(1-\gamma)\beta_I\frac{1}{\beta_\omega\omega_j^H + \beta_e e_j^H + \beta_I I_j^H} &= \lambda p_I && (12)\\ (1-a_1-a_2)U\pi_j\alpha_C(1-\gamma)\beta_I\frac{1}{\beta_\omega\omega_j^C + \beta_I I_j^C} &= \lambda p_I, && (13)\end{aligned}$$

where $\pi_i = \frac{q_i^\rho}{q_i^\rho + q_j^\rho}$ and $\pi_j = \frac{q_j^\rho}{q_i^\rho + q_j^\rho}$.

We first solve the optimal level of human capital investment. Specifically,

$\frac{(10)}{(12)}$:

$$\frac{\pi_i}{\pi_j}(\beta_\omega\omega_j^H + \beta_e e_j^H + \beta_I I_j^H) = \beta_\omega\omega_i^H + \beta_e e_i^H + \beta_I I_i^H; \tag{14}$$

$\frac{(10)}{(13)}$:

$$\frac{\alpha_H}{\alpha_C}\frac{\pi_i}{\pi_j}(\beta_\omega\omega_j^C + \beta_I I_j^C) = \beta_\omega\omega_i^H + \beta_e e_i^H + \beta_I I_i^H; \tag{15}$$

$\frac{(11)}{(12)}$:

$$\frac{\alpha_C}{\alpha_H}\frac{\pi_i}{\pi_j}(\beta_\omega \omega_j^H + \beta_e e_j^H + \beta_I I_j^H) = \beta_\omega \omega_i^C + \beta_I I_i^C; \tag{16}$$

$\frac{(11)}{(13)}$:

$$\frac{\pi_i}{\pi_j}(\beta_\omega \omega_j^C + \beta_I I_j^C) = \beta_\omega \omega_i^C + \beta_I I_i^C; \tag{17}$$

(15) + (17):

$$\frac{1}{\alpha_C}\frac{\pi_i}{\pi_j}(\beta_\omega \omega_j^C + \beta_I I_j^C) = \beta_\omega \sum_k \omega_i^k + \beta_I \sum_k I_i^k + \beta_e e_i^H; \tag{18}$$

$\frac{(14)}{(15)}$

$$\frac{1}{\alpha_C}\frac{\pi_i}{\pi_j}(\beta_\omega \omega_j^C + \beta_I I_j^C) = \beta_\omega \sum_k \omega_j^k + \beta_I \sum_k I_j^k + \beta_e e_j^H \tag{19}$$

(18) + (19):

$$I_j^{C*} = \frac{\alpha_C}{\beta_I}\pi_j W - \frac{\beta_\omega}{\beta_I}\omega_j^C \tag{20}$$

where $W = \beta_\omega \sum_k \sum_\iota \omega_\iota^k + \beta_I \sum_k \sum_\iota I_\iota^k + \beta_e \sum_\iota e_\iota^H$;

(14) + (16):

$$\frac{1}{\alpha_H}\frac{\pi_i}{\pi_j}(\beta_\omega \sum_k \omega_j^k + \beta_I \sum_k I_j^k + \beta_e e_j^H) = \beta_\omega \sum_k \omega_i^k + \beta_I \sum_k I_i^k + \beta_e e_i^H \tag{21}$$

$\frac{(17)}{(16)}$:

$$\frac{1}{\alpha_H}(\beta_\omega \omega_j^H + \beta_e e_j^H + \beta_I I_j^H) = \beta_\omega \sum_k \omega_j^k + \beta_I \sum_k I_j^k + \beta_e e_j^H \tag{22}$$

(21) + (22):

$$I_j^{H*} = \frac{\alpha_H}{\beta_I}\pi_j W - \frac{\beta_e}{\beta_I}e_j^H - \frac{\beta_\omega}{\beta_I}\omega_j^H; \tag{23}$$

By symmetry,

$$\begin{aligned} I_i^{H*} &= \frac{\alpha_H}{\beta_I}\pi_i W - \frac{\beta_e}{\beta_I}e_i^H - \frac{\beta_\omega}{\beta_I}\omega_i^H \\ I_i^{C*} &= \frac{\alpha_C}{\beta_I}\pi_i W - \frac{\beta_\omega}{\beta_I}\omega_i^C. \end{aligned}$$

We first note that

$$W = \beta_\omega \sum_k \sum_\iota \omega_\iota^k + \beta_e \sum_\iota e_\iota^H + \beta_I[(Y + w - c - wl)/p_I].$$

W measures the total inputs used to produce the children's human capital, which include health and cognitive endowments of both children, health shocks, and total investment allocated to children. The inputs are weighted by their relative importance in the production function.

We then consider π_ι, which measures the relative importance of child ι in the utility that parents derive from their children. First, $\pi_i + \pi_j = 1$. Second, the sign of $\partial \pi_i / \partial e_i^H$ is unambiguously determined by the degree of parental inequality aversion:[6]

$$sign(\frac{\partial \pi_i}{\partial e_i^H}) = sign(-\rho).$$

Following Becker and Tomes (1976), we interpret $d\pi_i/de_i^H$ as a *price effect*. As indicated in the first-order conditions of Equations (11) and (12), an early health shock on child i changes the *shadow prices* of the human capital investments on the child. If parents are sufficiently inequality averse ($\rho < 0$), then the price effect is positive. Otherwise, the price effect is negative. The sign of the cross sibling price effect, which is defined as $\partial \pi_i / \partial e_j^H$, is also unambiguously determined by ρ. However, the sign of the cross-sibling price effect is

---

[6] We derive that

$$\frac{\partial \pi_i}{\partial e_i^H} = \rho \frac{q_i q_j}{e_i^H} \frac{(\varepsilon_{q_i, e_i^H} - \varepsilon_{q_j, e_i^H})}{U^{2\rho}},$$

where ε_{q_i, e_i^H} is the elasticity of child i's quality with respect to an early health shock on himself, and ε_{q_j, e_i^H} is the elasticity of child i's quality with respect to an early health shock on his sibling. We assume that $\varepsilon_{q_i, e_i^H} - \varepsilon_{q_j, e_i^H} < 0$. This assumption derives from the parental utility function. If parents have symmetric preference, that is, $U(q_i, q_j) = U(q_j, q_i)$, then this assumption is automatically satisfied. The assumption of symmetric preference is also invoked in Behrman, Pollak, and Taubman (1982). We test this assumption in our empirical analysis, which is not rejected.

opposite to that of $\partial \pi_i / \partial e_i^H$:[7]

$$sign(\frac{\partial \pi_i}{\partial e_j^H}) = sign(\rho).$$

Therefore,

$$sign(\frac{\partial \pi_i}{\partial e_i^H} - \frac{\partial \pi_i}{\partial e_j^H}) = sign(-\rho).$$

The difference between self and cross sibling effects of early health shock on investments in the health and cognitive skills of child i are derived as follows:

$$\begin{aligned}
\frac{\partial I_i^{H*}}{\partial e_i^H} - \frac{\partial I_i^{H*}}{\partial e_j^H} &= \frac{\alpha_H}{\beta_I}(\frac{\partial \pi_i}{\partial e_i^H} - \frac{\partial \pi_i}{\partial e_j^H})W - \frac{\beta_e}{\beta_I}, \\
\frac{\partial I_i^{C*}}{\partial e_i^H} - \frac{\partial I_i^{C*}}{\partial e_j^H} &= \frac{\alpha_C}{\beta_I}(\frac{\partial \pi_i}{\partial e_i^H} - \frac{\partial \pi_i}{\partial e_j^H})W.
\end{aligned}$$

We derive the optimal consumption and labor supply and conduct comparative static analysis. By summing up Equations (11)-(14) and by using $\pi_i + \pi_j = 1$ and $\alpha_H + \alpha_C = 1$, we have

$$\frac{(1 - a_1 - a_2)U}{W}(1 - \gamma)\beta_I = \lambda. \tag{24}$$

By substituting Equation (25) into (9), we derive

$$c = \frac{a_1 W}{(1 - a_1 - a_2)(1 - \gamma)\beta_I}. \tag{25}$$

and $\frac{9}{10}$:

$$c = \frac{a_{1w}}{a_2} l. \tag{26}$$

By substituting Equation (27) into (26), we derive

$$\begin{aligned}
l^* &= \frac{a_2}{w} \frac{W}{\beta_I[1 - \gamma(1 - a_1 - a_2)]} \\
&= \frac{a_2}{w} \widetilde{W},
\end{aligned} \tag{27}$$

---

[7] The assumption of symmetry is a sufficient condition for this conclusion.

where $W = \beta_\omega \sum_k \sum_\iota \omega_\iota^k + \beta_e \sum_\iota e_\iota^H + \beta_I[(Y+w)/p_I)$ is the weighted family total resources adding the child endowments and early health shocks, and $\widetilde{W} = \frac{\overline{W}}{\beta_I[1-\gamma(1-a_1-a_2)]}$ is the weighted effective resources adjusted by the production technology. By substitute Equation (28) into (26), we have

$$c^* = a_1\widetilde{W}. \tag{28}$$

Equations (28) and (29) show that the demand for leisure and consumption has the standard form in the textbook when the utility function is a CD form. The optimal labor supply is

$$T^* = 1 - \frac{a_2}{w}\widetilde{W}.$$

The effect of early health shock on parental consumption is

$$\begin{aligned} \frac{\partial c^*}{\partial e_i^H} &= a_1(\frac{\partial \widetilde{W}}{\partial e_i^H}) \\ &= \frac{a_1\beta_e}{[1-\gamma(1-a_1-a_2)]} \\ &< 0, \end{aligned}$$

and the effect of early health shock on parental labor supply

$$\begin{aligned} \frac{\partial T^*}{\partial e_i^H} &= -\frac{a_2}{w}(\frac{\partial \widetilde{W}}{\partial e_i^H}) \\ &= -\frac{a_2\beta_e}{w[1-\gamma(1-a_1-a_2)]} \\ &> 0, \end{aligned}$$

as $\beta_e < 0$.

3 Appendix C: Discussion on Potential Measurement Errors in Constructing the Variable of Child Early Health Shocks

This section addresses potential concerns with measurement errors in constructing the variable of child early health shocks. The issue of measurement errors with twin-based estimator has been extensively discussed in the literature on returns to schooling. Griliches (1979) demonstrates that the within-twin estimator may exacerbate the problem of measurement errors with reported schooling years in estimating returns to education. A special issue of *Economics of Education Review* contains a recent exchange about twin-based estimators with measurement errors (Behrman and Rosenzweig, 1999; Bound and Solon, 1999; Neumark, 1999; Rouse, 1999). Strauss and Thomas (1998) discuss the measurement error problem with self-reported health status in developing countries.

We believe that the measurement error problem is not a major concern in our study. For illustration, we assume that child early health shocks, $e_{\iota,\tau}$, are not observed; instead, we observe

$$e^m_{\iota,\tau} = e_{\iota,\tau} + \nu_{\iota,\tau}. \tag{29}$$

We discuss three types of measurement errors. The first one is the classic measurement error such that $cor(\nu_{\iota,\tau}, e_{\iota,\tau}) = 0$, $cor(\nu_{\iota,\tau}, e^m_{\iota,\tau}) \neq 0$, and $cor(\nu_{\iota,\tau}, \epsilon^k_{i,\tau}) = 0$, where $\epsilon^k_{i,\tau}$ is the error term in the family investment equation (See Equation (8) in the paper).[8] This type of measurement errors can be addressed by factor methods. Specifically, we have asked different types of serious diseases in the questionnaire for each child. Therefore, we have multiple measures on child early health shocks. We estimate the measurement system and generate a factor score for early health shocks. Denote $m^r_{\iota,\tau}$ as the rth measurement of early

---

[8] The estimates are unbiased when $cor(\nu_{\iota,\tau}, e_{\iota,\tau}) \neq 0$, $cor(\nu_{\iota,\tau}, e^m_{\iota,\tau}) = 0$, and $cor(\nu_{\iota,\tau}, \epsilon^k_{i,\tau}) = 0$.

health shock on child ι ($\iota = i, j$) in family τ. We then write the measurement system as

$$m^r_{\iota,\tau} = \Psi^r_0 + \Psi^r_1 e_{\iota,\tau} + \varepsilon^r_{\iota,\tau}. \tag{30}$$

In this case, $e_{\iota,\tau}$ is a continuous variable and can be interpreted as the intensity of early health shocks. The classic measurement errors are no longer contained in $e_{\iota,\tau}$. The estimation results obtained by using the factor score are reported at the end of this section, which are qualitatively the same as those reported in the paper. Therefore, our results reported in the paper are robust to the presence of classic measurement errors.

The second type of measurement errors is that respondents may use different thresholds. Our measure of early health shocks is based on reported health histories. Some of the differences in the reported illness across households may simply reflect differences in the standard. This type of measurement errors is common with retrospectively constructed measures of health conditions (Strauss and Thomas, 1998; Smith, 2009). In this case,

$$\nu_{\iota,\tau} = \mu_\tau + \upsilon_{\iota,\tau}, \tag{31}$$

where μ_τ is the cross-household heterogeneity in reporting thresholds, and $\upsilon_{\iota,\tau}$ is the classic measurement errors such that $cor(\upsilon_{\iota,\tau}, \epsilon^k_{i,\tau}) = 0$.[9] If this is the case, $\nu_{\iota,\tau}$ may correlate with the family investment decision because it reflects the cross-household heterogeneity, which cannot be solved by factor methods. The twin design overcomes this problem. Parents have a same standard of reporting child illness between twin siblings. When we construct a dummy variable of early health shocks, the child who is reported to suffer from a serious disease is more likely to experience early health shocks than his or her twin sibling. Therefore, $cor(\nu_{\iota,\tau}, \epsilon^k_{i,\tau}) = 0$ after controlling for μ_τ in Equation (8) in the paper.

The third type includes the measurement error such that $cor(\nu_{\iota,\tau}, e^m_{\iota,\tau}) \neq 0$ and $cor(\nu_{\iota,\tau}, \epsilon^k_{i,\tau}) \neq 0$, which cannot be resolved by either the factor method or twin design. One source of this

---

[9] We assume that the cross-household heterogeneity in reporting thresholds, μ_τ, is linearly additive in the measurement equation above.

type of measurement errors comes from recall bias. Parents whose child who is currently sick may report that the child had also been sick in the past. In this case, the measurement error correlates with family investments and the measures of child early health shocks simultaneously. We believe that the recall bias is less of concern in our study because of three reasons. (*i*) Health history questions for each child were answered together by the father, mother, and children in our sample. (*ii*) The average age of children in our sample is 11. Given the young age of the twin, the recall period is not long. (*iii*) Parents and children are also asked to specify the timing and duration of each disease. This contextualization has the potential to increase recollection effort and further minimize recall error.

The third type of measurement errors may come from other sources besides the recall bias. For example, individual specific heterogeneity which is unobservable may exist in the investment equation. If the measurement error was not common for each twin pair but correlated with the unobservable heterogeneity, it correlated with the family investment decision. In this case, $cor(\nu_{\iota,\tau}, \epsilon^{k}_{i,\tau}) \neq 0$, and there exists a combined impact of endogenous variation and measurement errors, invalidating our twin design. If this is the case, the null hypothesis of the cross-equation restriction presented in Section 4 in the paper should rejected. The test results reported in Tables 3 and 4 in the paper fail to reject the null hypothesis. Therefore, we conclude that the combined impact of endogenous variation and measurement errors is not a major concern in our paper.[10]

---

[10] We thank an anonymous referee for pointing out the application of the cross-equation restriction to test the existence of endogenous measurement errors.

Table C1: Early Health Shocks and Family Investments (FE estimates)

	Dependent variable: Health investments (1)	Educational investments (2)
Early health shocks	0.306***	-0.056***
	[0.090]	[0.017]
Birth weight(kg): < 2	0.564***	-0.019
	[0.206]	[0.039]
Birth weight (kg): $2 - 2.5$	0.481***	0.016
	[0.164]	[0.031]
Birth weight (kg): $2.5 - 3$	0.390***	-0.009
	[0.140]	[0.027]
Male	0.088	-0.027
	[0.096]	[0.018]
# Pairs of twins	1,456	1,456

4 Appendix D: Child Early Health Shocks and Family Investments by Subsamples

This section discusses the effect of child early health shocks on family investments by subsamples. We divide the sample by *hukou* status (rural vs. urban), maternal education, household wealth, and the gender composition of twins. Before presenting the estimation results, Table D1 summarizes the family health and education investments by subsamples. We first note substantial differences between rural and urban groups. Both the health and education investments in the urban sample are about twice as much as those in the rural sample. These differences echo the big income gap between rural and urban areas in China. As far as maternal education is concerned, children of more educated mothers get much more investments in their health and education. Similarly, children from rich families get more investments in both health and education. In contrast, we do not find significant gender differences in family investments. The female twins get marginally more family investments than male twins do. But the family investment in the mixed sample with one male and one female twins is lower than the other two groups.[11]

Table D2 presents the estimates of the effects of child early health shocks on family investments by subsamples. We find significant differences in the compensating and reinforcing pattern across subsamples. First, the increase in health expenditures in favor of the sick twin in rural areas is not accompanied by a corresponding decrease in educational expenditure. This finding may be due to that families live in a subsistence level in rural areas, so that no further reductions in educational expenditures are possible.[12] In urban areas, instead, the amount of educational resources subtracted from the sick child almost exactly offsets, in monetary terms, the amount redistributed to pay for the medical expense.

---

[11] In the mixed sample, there is also no significant gender difference. The health investment on boys is RMB 167.79 and the education investment is RMB 797.48. For girls, they are RMB 156.18 and RMB 821.43, respectively.

[12] The family investment in children's education is RMB 632.48 in rural areas in the past 12 months prior to the survey, which is equivalent to USD 75.

Second, both the compensating health investment and the reinforcing education investment are more significant when mothers have a higher education level. Third, there is no significant difference in the compensating health investment behavior between poor and rich households. In contrast, the reinforcing education investment behavior is more pronounced in rich households that that in poor households. Finally, we find significant differences by gender. The compensating health investment and the reinforcing education investment are more pronounced in the female twin than male twin samples.[13]

---

[13] This may be due to the fact that the return to schooling is significant higher for females than that for males (Zhang, Zhao, Park, and Song, 2005).

Table D1: Summary Statistics on Family Health and Education Investments by Subsamples

	Health investments (RMB) (1)	Education investments (RMB) (2)
Whole sample		
	225.83	910.44
# Pairs of twins	1457	1456
Rural sample		
	149.80	632.48
# Pairs of twins	772	772
Urban sample		
	311.65	1224.17
# Pairs of twins	684	684
Less educated sample (maternal schooling years ≤ 9)		
	184.11	748.09
# Pairs of twins	1,109	1,109
More educated sample (maternal schooling years > 9)		
	359.18	1428.29
# Pairs of twins	347	347
Poor families (Family asset (score) $\leq$ sample mean)		
	167.15	568.49
# Pairs of twins	731	731
Rich families (Family asset (score) $>$ sample mean)		
	289.88	1254.07
# Pairs of twins	725	725
Male sample		
	238.65	909.13
# Pairs of twins	539	539
Female sample		
	254.54	977.63
# Pairs of twins	558	558
Mixed sample (one male and one female twins)		
	161.98	807.95
# Pairs of twins	359	359

Table D2: Child Early Health Shocks and Family Investments by Subsamples (FE estimates)

	Dependent variables	
	Health investments (1)	Education investments (2)
Whole sample		
Early health shocks	1.349***	-0.204***
	[0.243]	[0.047]
# Pairs of twins	1,456	1,456
Rural sample		
Early health shocks	1.548***	-0.054
	[0.351]	[0.047]
# Pairs of twins	772	772
Urban sample		
Early health shocks	1.148***	-0.320***
	[0.343]	[0.079]
# Pairs of twins	684	684
Lowly educated sample (maternal schooling years ≤ 9)		
Early health shocks	1.182***	-0.089*
	[0.279]	[0.048]
# Pairs of twins	1,109	1,109
Highly educated sample (maternal schooling years > 9)		
Early health shocks	1.672***	-0.472***
	[0.502]	[0.115]
# Pairs of twins	347	347
Poor families (Family asset (score) $\leq$ sample mean)		
Early health shocks	1.433***	-0.120***
	[0.371]	[0.043]
# Pairs of twins	731	731
Poor families (Family asset (score) $>$ sample mean)		
Early health shocks	1.260***	-0.280***
	[0.323]	[0.081]
# Pairs of twins	725	725

Table D2: Child Early Health Shocks and Family Investments by Subsamples (FE estimates) (Cont.)

	Dependent variables	
	Health investments (1)	Education investments (2)
Male sample		
Early health shocks	1.084***	-0.171***
	[0.380]	[0.052]
# Pairs of twins	539	539
Female sample		
Early health shocks	2.078***	-0.410***
	[0.421]	[0.089]
# Pairs of twins	558	558
Mixed sample (one male and one female twins)		
Early health shocks	0.845*	-0.006
	[0.494]	[0.114]
# Pairs of twins	359	359

Source: CCTS. Notes: Each entry comes from a separate regression. Robust standard errors are in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Birth weight is controlled for in each regression; gender has been controlled for in estimations except the male and female samples; rural has been controlled for in estimations except the rural and urban samples.

5 Appendix E: Early Health Shocks and Child Human Capital (full results)

Table D1: Early Health Shocks and Child Health

	Dependent variables: Height *z*-score (1)	 Weight *z*-score (2)	 BMI *z*-score (3)	 Health Status (4)
	2SLS estimates			
Early health shocks	-0.1	-0.422***	-0.395***	-0.513***
	[0.130]	[0.118]	[0.151]	[0.078]
Health investments	0.07	0.118**	0.160**	0.047
	[0.064]	[0.059]	[0.077]	[0.039]
Birth weight (kg): < 2	-0.356***	-0.516***	-0.360***	-0.043
	[0.088]	[0.080]	[0.106]	[0.053]
Birth weight (kg): $2 - 2.5$	-0.269***	-0.341***	-0.201**	-0.014
	[0.071]	[0.065]	[0.086]	[0.043]
Birth weight (kg): $2.5 - 3$	-0.199***	-0.206***	-0.08	-0.001
	[0.061]	[0.056]	[0.074]	[0.037]
Male	0.03	0.024	-0.009	-0.021
	[0.038]	[0.035]	[0.046]	[0.023]
	Reduced-form estimates			
Early health shocks	-0.004	-0.263***	-0.201*	-0.449***
	[0.096]	[0.086]	[0.113]	[0.057]
Birth weight (kg): < 2	-0.317***	-0.454***	-0.275***	-0.018
	[0.081]	[0.072]	[0.094]	[0.048]
Birth weight (kg): $2 - 2.5$	-0.235***	-0.284***	-0.125*	0.009
	[0.064]	[0.057]	[0.075]	[0.038]
Birth weight (kg): $2.5 - 3$	-0.170***	-0.156***	-0.012	0.019
	[0.055]	[0.049]	[0.064]	[0.033]
Male	0.034	0.032	0.003	-0.018
	[0.038]	[0.034]	[0.044]	[0.022]
# Pairs of twins	1,418	1,430	1,408	1,450

Table D2: Early Health Shocks and Child Education (1)

	Dependent variables:			
	Literature		Mathematics	
	score	relative measure	score	relative measure
	(1)	(2)	(3)	(4)
	2SLS estimates			
Early health shocks	-3.990*	-0.168	-4.697*	-0.496***
	[2.045]	[0.145]	[2.432]	[0.157]
Education investments	6.124	0.904**	3.598	0.17
	[6.313]	[0.426]	[7.459]	[0.466]
Birth weight (kg): < 2	-2.845**	-0.136	-2.336	-0.095
	[1.395]	[0.096]	[1.695]	[0.104]
Birth weight (kg): $2 - 2.5$	-1.157	-0.091	-0.424	-0.023
	[1.093]	[0.077]	[1.316]	[0.083]
Birth weight (kg): $2.5 - 3$	-1.255	-0.055	-0.128	0.02
	[0.929]	[0.066]	[1.121]	[0.071]
Male	-2.755***	-0.150***	-0.609	-0.019
	[0.663]	[0.046]	[0.791]	[0.050]
	Reduced-form estimates			
Early health shocks	-5.142***	-0.352***	-5.372***	-0.531***
	[1.665]	[0.110]	[1.996]	[0.127]
Birth weight (kg): < 2	-3.041**	-0.149	-2.482	-0.098
	[1.381]	[0.091]	[1.674]	[0.105]
Birth weight (kg): $2 - 2.5$	-1.07	-0.076	-0.408	-0.021
	[1.090]	[0.073]	[1.320]	[0.084]
Birth weight (kg): $2.5 - 3$	-1.318	-0.066	-0.166	0.018
	[0.927]	[0.062]	[1.122]	[0.071]
Male	-2.903***	-0.173***	-0.68	-0.023
	[0.645]	[0.043]	[0.780]	[0.049]
# Pairs of twins	1,355	1,426	1,332	1,420

Table D2: Early Health Shocks and Child Education (2)

	Dependent variables: Good Student Awards	Awards in Contests	Grade Repetition	Doing minor actions in class
	(1)	(2)	(3)	(4)
	2SLS estimates			
Early health shocks	-0.199***	-0.067	0.025	0.296**
	[0.072]	[0.042]	[0.038]	[0.121]
Health investments	0.075	0.103	-0.235**	-0.639*
	[0.216]	[0.126]	[0.113]	[0.387]
Birth weight (kg): < 2	-0.092*	0.007	-0.008	-0.067
	[0.048]	[0.028]	[0.025]	[0.087]
Birth weight (kg): $2 - 2.5$	-0.048	0.005	0.004	-0.098
	[0.038]	[0.022]	[0.020]	[0.069]
Birth weight (kg): $2.5 - 3$	-0.019	0.001	0	0.01
	[0.033]	[0.019]	[0.017]	[0.059]
Male	-0.129***	-0.029**	-0.002	0.256***
	[0.023]	[0.013]	[0.012]	[0.041]
	Reduced-form estimates			
Early health shocks	-0.215***	-0.088***	0.073**	0.396***
	[0.058]	[0.033]	[0.029]	[0.101]
Birth weight (kg): < 2	-0.093*	0.006	-0.005	-0.056
	[0.049]	[0.028]	[0.024]	[0.084]
Birth weight (kg): $2 - 2.5$	-0.047	0.007	0.001	-0.108
	[0.039]	[0.022]	[0.019]	[0.067]
Birth weight (kg): $2.5 - 3$	-0.02	-0.001	0.003	0.018
	[0.033]	[0.019]	[0.017]	[0.057]
Male	-0.131***	-0.032**	0.004	0.272***
	[0.023]	[0.013]	[0.011]	[0.039]
# Pairs of twins	1,456	1,456	1,456	1,440

Table D3: Early Health Shocks and Child Socioemotional Skills

	Dependent variables:			
	Feel lonely	Easily distracted	Easily frightened	Emotional instable
	(1)	(2)	(3)	(4)
	2SLS estimates			
Early health shocks	0.132***	0.121**	0.066	0.085***
	[0.041]	[0.056]	[0.046]	[0.024]
Education investments	-0.158	-0.15	-0.383***	-0.113
	[0.123]	[0.166]	[0.139]	[0.073]
Birth weight (kg): < 2	0	-0.009	-0.016	0.018
	[0.027]	[0.037]	[0.031]	[0.016]
Birth weight (kg): $2 - 2.5$	0.016	-0.02	0.013	0.010
	[0.022]	[0.030]	[0.025]	[0.013]
Birth weight (kg): $2.5 - 3$	0.007	-0.027	-0.015	0.009
	[0.019]	[0.025]	[0.021]	[0.011]
Male	-0.004	0.071***	-0.016	0.007
	[0.013]	[0.018]	[0.015]	[0.008]
	Reduced-form estimates			
Early health shocks	0.165***	0.151***	0.144***	0.108***
	[0.032]	[0.044]	[0.033]	[0.019]
Birth weight (kg): < 2	0.002	-0.007	-0.011	0.02
	[0.027]	[0.037]	[0.027]	[0.016]
Birth weight (kg): $2 - 2.5$	0.013	-0.023	0.007	0.009
	[0.022]	[0.029]	[0.022]	[0.013]
Birth weight (kg): $2.5 - 3$	0.009	-0.025	-0.011	0.011
	[0.018]	[0.025]	[0.019]	[0.011]
Male	0	0.075***	-0.006	0.01
	[0.013]	[0.017]	[0.013]	[0.007]
# Pairs of twins	1,456	1,456	1,456	1,450

References

Behrman, J., and M. Rosenzweig (1999): "Ability biases in schooling returns and twins: a test and new estimates," *Economics of Education Review*, 18(2), 159–167.

Bound, J., and G. Solon (1999): "Double trouble: on the value of twins-based estimation of the return to schooling," *Economics of Education Review*, 18(2), 169–182.

Chen, L. C., and N. S. Scrimshaw (1983): *Diarrhea and malnutrition: interactions, mechanisms, and interventions.* Plenum Press.

Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013): "Evidence on the impact of sustained exposure to air pollution on life expectancy from Chinas Huai River policy," *Proceedings of the National Academy of Sciences*, 110(32), 12936–12941.

Ebenstein, A. (2012): "The consequences of industrialization: evidence from water pollution and digestive cancers in China," *Review of Economics and Statistics*, 94(1), 186–201.

Fang, Z.-Y., et al. (2005): "Sentinel hospital surveillance for rotavirus diarrhea in the People's Republic of China, August 2001–July 2003," *Journal of Infectious Diseases*, 192(Supplement 1), S94–S99.

Griliches, Z. (1979): "Sibling Models and Data in Economics: Beginnings of a Survey," *Journal of Political Economy*, 87(5), S37–S64.

Neumark, D. (1999): "Biases in twin estimates of the return to schooling," *Economics of Education Review*, 18(2), 143–148.

Rouse, C. (1999): "Further estimates of the economic return to schooling from a new sample of twins," *Economics of Education Review*, 18(2), 149–157.

Smith, J. (2009): "The impact of childhood health on adult labor market outcomes," *The Review of Economics and Statistics*, 91(3), 478–489.

Strauss, J., and D. Thomas (1998): "Health, nutrition, and economic development," *Journal of Economic Literature*, 36(2), 766–817.

Zhang, J., Y. Zhao, A. Park, and X. Song (2005): "Economic returns to schooling in urban China, 1988 to 2001," *Journal of Comparative Economics*, 33(4), 730–752.