Piped water accessibility and child malnutrition: Evidence from developing countries

Dao Nguyen Dinh[1, 2, *]; Luke Connelly[2, 4]; Stephen Birch[2]; Ha Nguyen[3]

August 2022

Preliminary draft, please do not circulate

Abstract

Developing countries have vigorously implemented policies to promote safe drinking water coverage. The provision of piped clean water on the premises, especially for the poor, has not, however, been a strong focus of recent research. This study aims to estimate the impact of access to on-premises piped water on the malnourished health of children in three developing countries, Ethiopia, India, and Vietnam. The study uses panel data that were collected through Young Lives studies (YLS) since 2000 to examine the impact on child malnutrition in developing countries by switching from other water sources to safe piped water. The data also supports focusing on around 7100 children between ages 5 and 15 who stayed in the same community during all surveyed rounds. The YLS panel also enables us to address an important endogeneity problem that has afflicted previous studies based on cross-sectional data. The approach includes the use of two-way fixed-effects estimators and instrumental variables. The results show that access to piped water sources inside the dwelling insignificantly affect the height and BMI of children. However, if children of educated women had access to piped water at home, their height will be much greater than that of their peers. The findings suggest that the governments should pay more attention to expanding piped water supply in the yard for the whole population, particularly having more specific support for the poor. Furthermore, enhancing the health knowledge of safe water for mothers may be an essential part of policies that are designed to promote the beneficial effects of safe water and sanitation in developing countries.

Key words: child health; piped water accessibility, Young Lives Study; fixed effects; endogeneity; leave-out strategy

JEL-codes: D04; J13; I15, I38.

---

[1] School of Economics, The University of Queensland, St Lucia, Queensland, Australia
[2] Centre for the Business and Economics of Health, The University of Queensland, St Lucia, Queensland, Australia
[3] Telethon Kids Institute, Perth, Australia
[4] Department of Sociology and Business Law, The University of Bologna, Bologna, Italy
* Corresponding author: d.nguyendinh@uq.edu.au.

1. Introduction

Due to its long-lasting effect on child development, child malnutrition has garnered much attention (Black *et al.*, 2008; Ali, 2019). Wells *et al.* (2020) highlighted its life-course health hazard, particularly serious if severe malnutrition occurs in early life. Globally, children suffer a higher risk of disease and death due to malnutrition (WHO, 2020). Malnutrition in children is substantially more hazardous than other common health conditions such as diarrhea, malaria, and pneumonia (WHO, 2019).

Extremely persistent and untreated childhood malnutrition may increase the likelihood of physical and mental issues in adulthood (Ricciuti, 1981). Stunting, a consequence of continuous undernutrition, is significantly connected with long-term effects such as impaired health, delayed cognitive development, economic output, and reproductive capacity (Dewey and Begum, 2011). Some data also suggest a link between stunting, an increased risk of noncommunicable diseases, and childhood obesity (Schott et al., 2019). In the meantime, cardiovascular disease, coronary heart disease, and death rate in mature age may be a result of the early-life obesity problem (Schott et al., 2019). and untreated childhood malnutrition may increase the likelihood of physical and mental issues in adulthood (Ricciuti, 1981). Stunting, a consequence of persistent undernutrition, is significantly associated with long-term outcomes, including impaired health, cognitive development, economic productivity, and potential reproduction (Dewey and Begum, 2011). Some evidence also indicates the association between stunting, a higher risk of non-communicable diseases, and being overweight as children grow up (Schott *et al.*, 2019).

Determining the primary causes of child malnutrition will facilitate the development of strategies to mitigate its long-term effects. The Millennium Development Goals (MDGs) emphasized the importance of a safe water supply in preventing disease, particularly in less-developed nations. The Sustainable Development Goals (SDGs) reference the ongoing discussion regarding the significance of safe water access to all elements of sustainable development in order to ensure that *"no one will be left behind."* UNICEF's (1990) framework acknowledged the importance of safe water access to child nutrition. Consequently, access to safe water is a crucial proxy for health services and the environment, one of three underlying causes of children's nutrition.

The impact of safe water accessibility on child malnutrition could be explained by reducing the risk of diarrhea (Mills and Cumming, 2016; van Cooten *et al*., 2019). It is well-known that germs and worms from contaminated water sources frequently induce diarrhea (Conant and Fadem, 2008). Consequently, the child's ability to eat and absorb food nutrients may be greatly impaired, leading to malnutrition (Dewey and Mayers, 2011; van Cooten *et al*., 2019).

A reduction in diarrhoeal disease may reduce the risk of child malnutrition. The probability of diarrhea is reduced by 23% if households switch from unsafe water sources to piped water on-premises (WHO, 2014). The diarrhea-health reduction will be even more pronounced if children use clean water after using effective treatments such as boiling, filtering, or safe storage (WHO, 2014; Jalan and Ravallion, 2003). Using water filters or high-quality piped water effectively reduces child diarrhea (Mills and Cumming, 2016). Besides, the simultaneous presence of safe water accessibility and sanitation facilities might significantly reduce the probability of diarrhea in children (Cairncross *et al.*, 2010).

The current literature has mainly focused on the impact of safe water access on child health at a specific stage, e.g., children under five years old, school-age children, or adolescents separately (Jalan and Ravallion, 2003; Lamichhane and Mangyo, 2011; Devoto *et al.*, 2012; van Cooten *et al.*, 2019). This leads to a neglected area in dynamic research throughout the development from birth to adolescence of each individual in switching water supply resources that cannot be identified in cross-sectional analysis (Mangyo, 2008; Headey and Palloni, 2019).

Moreover, cross-sectional studies typically ignore the presence of unobserved time-invariant heterogeneity, subsequently leading to an endogeneity problem (Woolwridge, 2010; Woolwridge, 2015; Le and Nguyen, 2018; Leszczensky and Wolbring, 2019). Existing studies also confirmed the endogeneity nature of water accessibility in modeling health outcomes (Mangyo, 2008; Lamichhane and Mangyo, 2011; Headey and Palloni, 2019), resulting in inconsistent and bias results in cross-sectional estimations.

Methods based on panel data have gained prominence for tackling the endogeneity problem. Mangyo (2008) used community-level changes in water access as the instrumental variable of an individual's in-yard water accessibility and the dynamic panel data model to estimate the impact of water access on child anthropometrics. The author also denoted that the self-selection bias of water projects at the community level could underestimate the estimated results, leaving the limitation of the research. Headey and Palloni (2019) used difference-in-difference regression with controlling the regional fixed effects to address the endogeneity problem of the treatment variable. However, the omitted time-variant variables could inconsistently affect their estimate.

Developing countries have vigorously implemented policy concerns to promote safe drinking water coverage. For example, in the early 2000s, Vietnam issued the National Rural Clean Water Supply and Sanitation Strategy to improve people's access to safe drinking water. Since 2006, the policy has been integrated as a part of The National Targeted Programme for New Rural Development and The National Targeted Programme on Sustainable Poverty Reduction (UNICEF, 2020). Also, with great ambition to achieve the MDGs in national access to clean drinking water, the primary programmes in Ethiopia can be mentioned as the Ethiopian Water Resources Management Policy in 1999[1], Growth and Transformation Plan I (GTP I) in 2010, and GTP II in 2015 (UNICEF, 2018). Besides, India's government also considers the provision of safe drinking water as the top priority of the National Water Policy (NWP), initially imposed in 1987 and revised in 2002 and 2012 (Sharma, 2017).

Their programs and policies, however, do not emphasize the provision of piped clean water on the premises. Although the safe water coverage in these countries has quickly increased thanks to these implemented policies and programmes, the percentage of households with piped water in the yard has been low, especially in rural areas and among the poorest people (ADB, 2010; World Bank, 2014; UNICEF, 2018). Most households have access to clean drinking water outside their houses (ADB, 2010). Sharma (2017) also highlighted that India's NWP had focused more on monitoring and managing surface and underground water sources.

The piped water in the yard can be the best quality compared to the clean water outside the home (Mangyo, 2008). The piped water source inside the home could help reduce water storage and collection time, thus reducing the risk of contamination and germ growth (Jalan and

---

[1] More information: http://www.fao.org/faolex/results/details/en/c/LEX-FAOC158196/

Ravallion, 2003; Mangyo, 2008). Contamination can occur by touching hands or utensils directly to water in containers or tanks, especially when these containers and tanks are not covered (Overbo *et al.,* 2016). Its better quality subsequently positively impacts children's health (Mangyo, 2008). Besides, households with piped water access on the premises will help mothers reduce the time to fetch water and spend more time taking care of their children (Mangyo, 2008; Lamichhane and Mangyo, 2011).

The work's objective is to estimate the impact of access to on-premises piped water, the essential and safe drinking water sources, on child health in terms of malnutrition status in three developing countries, Ethiopia, India, and Vietnam. Based on the theories of Grossman (1972) and Jacobson (2000), we argue that access to in-yard piped water, an environmental factor, might affect child health. More specifically, consuming in-yard piped water could negatively correlate with health deterioration, leading to improved health outcomes. Therefore, we aim to examine the question, "*Do improvements in access to in-yard piped drinking water lead to improvements in health outcomes, child malnutrition*?". Besides, we also specifically look at the following question: *What is the role of maternal education on the health effect of access to piped water sources in the yard?*

In two significant ways, this work contributes to the body of knowledge. First, by utilising the most recent panel data collected as part of the Young Lives Study (YLS) since 2000, this study is able to overcome the drawbacks of prior studies that relied on cross-sectional data. The panel YLS data allows us to see the impact of switching to safe piped water on child malnutrition. Besides, we propose to use fixed-effects (FE) and fixed-effects instrumental variable models (FEIV) for analysis. The suggested methodology will produce a reliable estimate since it addresses the endogeneity problem of water accessibility caused by omitted variable bias, adverse causality, and measurement error in variables. The policy implications will be reasonable with consistent and unbiased estimates. Second, when we examine the causal relationship between access to piped water and child health in developing nations, our research can promote the goal of the Sustainable Development Goals that *"no one will be left behind."* In addition, our recommended data, YLS, oversampled children from poor communities, a disadvantaged demographic in any country. We can focus our attention on the role of clean drinking water access in the health of children from poor groups over a long research period.

We follow the strategy of using the leave-out community ratio as an internal IV to overcome the endogeneity problem of water accessibility in child health. Mangyo (2008) and Lamichhane and Mangyo (2011) used the data on the disease symptoms and concluded that the leave-out strategy indicator could be an invalid IV because of the endogenous water project placement in China and Nepal. As a result, their estimate could be inaccurate. We follow their strategy and see that the correlations between the treatment variable and the incidence of water-irrelevant and water-relevant disease symptoms are not statistically significant. We can conclude that the water project placement could be exogenous across communities. However, the results of the Hausman test from the FEIV estimator suggest that the treatment variable, piped water accessibility in the home, could be exogenous. The FE method, therefore, might be appropriate to estimate the causal effect of having a piped water supply on the premises on the health of children.

We found that children with on-premises piped water supply will have higher BMI and height indicators than their peers. However, our findings are not statistically significant. Moreover, we find that children born to mothers with a higher level of education are more likely to have access to piped water sources than those born to less-educated mothers. If children from educated mothers have access to piped water in their families, they will have a significantly higher height than their peers. Our findings support the argument of Mangyo (2008) that access to piped water sources in home and maternal education may serve as complements in producing the heath of children.

The following section presents our study's basic health economic theory and established model specification. Section 3 will describe the data and descriptive statistics for all selected variables. The estimated results will be discussed in Section 4, followed by Section 5 of Robustness check and Section 6 of Heterogeneity. Section 7 concludes.

2. Empirical framework

Based on the time allocation theory of Becker (1965), Grossman (1972) built his model of health demand. In his model, individual health might be seen as his/her capital good. Extending the basic Grossman's model, Jacobson (2000) considered each family as a unit and then proposed three primary models, including the single individual model, the husband-wife model, and the parent-child model. According to her parent-child model, child health demand might be constructed based on their family utility with respect to their parental wealth condition, the time allocation constraint, and the health production constraint of their parents and their own. Two of these models are our motivation.

Accordingly, each household wants to maximize its lifetime utility, which is constructed as follows:

$$\int_{t=0}^{T} e^{-\rho t} U(s_t(H_t), Z_t, TL_t)$$

Accordingly, the lifetime utility of each family depends on the sick time of the children (s_t), the total consumption of other goods and services (Z_t), and leisure time (TL_t), as measured at time *t*. The time discount factor is denoted as ρ. Mathematically, the intertemporal utility has the properties of strictly convex, a non-decreasing nature for F_t, Z_t, and a non-increasing nature for s_t. s_t is endogenous and non-increasing in H_t, child health capital ($\partial s_t / \partial H_t < 0$), as follows: $s_t = \alpha_1 H_t^{-\alpha_2}$, where $\alpha_i > 0$.

Following Grossman (1972) and Jacobson (2000), the utility will be constrained by the health production of children (H_t) which is positively affected by the investment of parents in health capital (I_t) and negatively affected by the health deterioration (δ_t). Health deterioration is the function of environmental factors (X_t). Remarkably, the unhealthy environmental condition will increase at such a rate ($\partial \delta_t / \partial X_t > 0$) and vice versa.

$$\dot{H}_t = I_t - \delta(t, X_t) * H_t$$

Besides, each family will face its asset accumulation constraint,

$$\dot{A} = rA_t + Y(s_t) - \pi_t^I I_t - \pi_t^Z Z_t$$

and the time allocation constraint,

$$TW_t + TM_t + TZ_t + TL_t + s_t = 365$$

and the endpoint conditions,

H_0 and A_0 are given; $H_T = H_{min};\ A_T \geq 0$

Where π_t^I and π_t^Z are marginal costs of investment and consumption of other goods and services, respectively. The stock of assets will vary overtimes by the difference between the total income $(rA_t + Y_t)$ and the total expenditure on investment and consumption. $Y(s_t)$ is the parental wage income, which is produced by sick time (s_t) of their child $(\partial Y_t/\partial s_t < 0)$. r is the return rate on asset capital. TW_t is working time. TM_t and TZ_t are parental time on childcare and time uses on food and other commodities consumption, in turn.

Through the Hamiltonian function and its first-order conditions, the optimal health demand function for a child could be obtained (see section S1 in the *Supplementary material file* for more details). The optimal demand function (H_t^*) points out that child health will be affected by the wage rate, the marginal cost of investment (which depends on M_t-medical care, and TM_t), the health stock depreciation rate (which is related to time and X_t), and a set of exogenous variables.

$$H_t^* = H(w_t, \pi_t^I, \delta(t, X_t), (\,r, \alpha_1, \alpha_2))$$

Based on the optimal health demand function, we build the model to estimate the impact of water accessibility on child health as below:

$$y_{it} = x_{it}\beta + WA_{it}\pi + c_i + \epsilon_{it} \qquad \text{(Eq. 01)}$$

Where, y_{it} is the dependent variable, child health. x_{it} is a set of independent variables, including intercept, a set of time dummies, a set of strictly exogenous variables, which are strictly exogenous with respect to the idiosyncratic error term ϵ_{it}. c_i is a set of time-invarying household-specific unobservable variables. WA_{it} is a dummy reflecting whether household access to piped water on the premises or not. WA_{it} is a proxy to X_t in the optimal health demand function.

Endogeneity is the most critical challenge to estimate the impacts of improved water accessibility on child health. A household's safe water accessibility could be endogenous due to its relationship with some unobservable factors that also affect child health (Mangyo, 2008). For example, women empowerment in developing countries, unobserved in the data, could be positively correlated to both the probability of access to piped water and the child health status of poor households. The participation and the voice of women in safe-water projects at their community levels, could support increasing their household's access to improved water sources (Ray, 2007; Allen *et al.*, 2018). The second reason of endogeneity is reverse causation. Poor access to drinking water possibly brings many adverse impacts on child health. And conversely, the higher investment in on-premises piped water supply could happen in households with bad-health children (Mangyo, 2008). Moreover, measurement error in health indicators can be happened with the self-reported household survey. Consequently, endogeneity might lead to inconsistent estimates, which could not account for causality in the population. We will use two-way fixed effects with instrumental variables (FEIV) to overcome the problem.

In Eq.01, the panel data will support eliminating the effect of time-invariant unobserved variables (c_i). However, the estimate of the endogenous treatment variable can still be biased due to the possible correlation between time-variant unobservable characteristics with the treatment regressor and the response outcome. Therefore, the FEIV approach is to overcome that endogeneity problem in the treatment variable. Notably, we need an auxiliary equation for the endogenous variable, piped water accessibility, with the presence of instrumental variables (IVs) as follows:

$$WA_{it} = x_{it}\tau + Z_{it}\sigma + \beth_{it} \qquad \text{(Eq. 02)}$$

Where Z_{it} is a set of instruments. $\beth_{it}$ denotes the idiosyncratic error term. According to Wooldrige (2012), we need to set two conditions for IVs. The first condition is the identification condition which assumes a sufficient correlation between Z_{it} and WA_{it}. Second, instrumental variables should be exogenous, meaning that they are uncorrelated with error terms ($\beth_{it}$).

3. Data and variables

The study use data from the Young Lives Studies project (YLS), which has been conducted to collect data by the University of Oxford in four countries, including Ethiopia, India, Peru, and Vietnam. YLS survey interviews the same individual repeatedly over time. Five waves are commencing in 2002, with two cohorts: Young Cohort (YC) with 2000 children born in 2001 and 2002; and Old Cohort (OC) with 1000 children born in 1994 and 1995 (excluding Peru with only 700 children in the OC). The oversampling survey of children from low-income areas is consequently not nationally representative. The estimated results in following sections will be interpreted for children living in the poor communities of these selected countries. In terms of sampling design, the YLS conducted multistage sampling: randomly selected sites and then randomly selected enumeration areas (communes or villages). The sites are selected from only Andhra Pradesh in India, while others are chosen randomly across the nation. We will analyse cohort data from Vietnam, India, and Ethiopia after excluding data from Peru, as Peruvians have universal access to piped water inside the premises. Our data allows for studying the subject during the most critical period of early childhood development, from birth to age 15.

The malnutrition outcome in this research will be related to the child's BMI and height. These measurements were adjusted by Z-scores to compare each anthropometric indicator with the World Health Organisation's growth standards in terms of child age. Mainly, the study uses BMI-for-age Z-score and height-for-age Z-score, which was calculated using WHO AnthroPlus, for analysis (WHO, 2009). These outcome indicators are also widely used in previous studies to estimate the child health impacts of water accessibility, namely Mangyo (2008); Lamichhane and Mangyo (2011). In accordance with Mangyo (2008), the sample is restricted to those having absolute values of these health indicators less than four to reduce the risk of height and/or weight misreporting. Besides, rounds 4-5 in the OC's data are also excluded because of missing values of anthropometric indicators. Furthermore, we will employ an alternative measure of anthropometric indicators as dummies in a robustness check.

A dummy reflecting whether a household has piped water supply in their dwelling or yard is chosen as a primary regressor. The data for Round 1 will be excluded because the question in piped-water accessibility is not consistent with that in other waves. However, as discussed above, piped-water accessibility could be endogenous due to its correlation with unobservable variables or its reverse causal effect on the response variable, the child's health. Mangyo (2008) and Lamichhane and Mangyo (2011) used the leave-out community ratio as an internal IV to overcome the endogeneity problem of water accessibility in child health. This ratio reflects the percentage of households in a commune having an in-yard piped-water source, not including the answer reported by household j itself. Accordingly, the probability of installing piped water sources in the yard of a family is highly correlated with community average behaviours, a phenomenon widely referred to as the causal peer effect, introduced by Maski (1993).

This leave-out strategy would reduce the endogeneity of household water access by taking advantage of differences in community-level water supply infrastructure supported by governments or NGO projects rather than their individual demands (Lamichhane and Mangyo, 2011). Households' self-selection of access to piped water sources could be eliminated by using the leave-out strategy (Lamichhane and Mangyo, 2011). The study, thus, creates this ratio at the commune level as the internal IV of in-yard water access. In order to construct this non-self-community instrument, the total number of households within each commune-round having in-dwell piped water sources is calculated, excluding the concerned household. Then, the leave-out instrument is the ratio of that number with the total number of households within each commune round.

Based on data availability and the theoretical framework, the study will select other covariates to control the child's, parent's, household's, and commune's characteristics. The local population size is collected from the commune data. The real 2006 price defines the value of food expenditure per capita after converting to USA dollars. We also use the consumer durables index to proxy a household's wealth. Finally, we restrict the data to only children who attended all rounds of the survey and lived in the same community for the duration of the study. In order to aggregate the panel data, the study will employ a sample of around 900 and 1460 children in OC and YC, respectively. In particular, the observations for each round were compiled in each selected country.

The summary statistic for the estimated sample of the study is reported in Tables 1-2. Accordingly, the percentage of YL households using piped water in their yards or dwellings tends to increase. In 2002, around 9% of Vietnamese households in the Young Cohort data accessed in-yard piped water which increased to 27% after 14 years. The figures for Vietnamese households in Older Cohort are 12% and 18% in 2002 and 2006, respectively. The same trend is found in India's and Ethiopia's samples. Besides, the leave-out indicator in piped water access increased over time in all countries.

Table 2 presents the compare-means tests of these selected variables between two household groups with and without using in-yard piped water. We found significant and negative differences in the mean of health indicators between the two groups. Children living in houses with piped water have significantly higher anthropometric indicators than their peers without piped water. Furthermore, Figures 1-3 also show such differences in the Kernel density measure. Accordingly, we also see that variations in child health outcomes approximate the normal distribution. However, it is essential to note that the relationship between child anthropometric health and in-yard piped-water access may be affected by omitted variables, adverse causality, and variable-measurement error. Using proposed econometrics methods, the concern might be resolved.

4. Estimated results

4.1. On-premises piped water accessibility and child anthropometric outcomes

To estimate the effect of access to on-premises piped water on child anthropometric outcomes, we perform a multi-level regression analysis for Eq.01, built-in Section "*Empirical framework*". The results are reported in Table 3. The BMI-for-age Z-score and height-for-age

Z-score are used as response variables. Three regression approaches, namely the Pooled Ordinary Least Square (Pooled OLS), Fixed-Effects (FE), and FEIV, are separately computed for each outcome measure. Mainly, columns 1 and 4 reflected the results of Pooled OLS in each health outcome. Estimated results from the FE method will be shown in columns 2 and 5, while estimates from the FEIV method will be shown in columns 3 and 6. Standard errors are clustered at the individual level.

Compared to Pooled OLS, FE methods provide smaller estimated coefficients of piped water accessibility in magnitude. The effects of time-invariant observables and unobserved variables are directly eliminated by utilizing fixed effects (Le and Nguyen, 2018). Therefore, FE is preferable to Pooled OLS for reducing the endogeneity problem by eliminating the influence of time-invariant unobserved confounders. However, as discussed above, the endogeneity problem of water accessibility in modeling health outcomes could still happen with the presence of time-variant omitted variables, reverse causality, or error measurement. Using IVs is important to cope with the endogeneity problem caused by those reasons.

The FEIV method uses the leave-out strategy in piped water as a single IV of the endogenous variable, water accessibility. First, the first-stage regression results in Table 3 (see Columns 3 and 6) suggest that the selected IV significantly affects households’ ability to have improved on-premises piped water sources. The average proportion of a type of water source in a community influence positively the likelihood of a household obtaining piped water. It indicates that a household residing in a community with a greater rate of piped water use will be more likely to install this sort of drinking water.

Second, the F tests for excluded instruments under the Kleibergen-Paap and Wald statistics for all models exceed 10, the rule-of-thumb value (Stock and Yogo, 2005), hence rejecting the null hypothesis of a weak instrument. Regarding the Hausman test, the results indicate that the treatment variable, piped water accessibility, is exogenous in both models. Therefore, the FE method is appropriate to estimate the impacts of piped water accessibility on child anthropometric outcomes.

The results of the FE estimator in Table 3 present that the estimated impacts of piped water accessibility on child height and BMI are positive. It means that children with in-yard piped water sources will have higher indicators of height and BMI than their peers, on average. The FE specification in column 5 reveals that children with on-premises piped water intervention have an average BMI that is 0.019 points higher than children without this water supply. However, our findings are not statistically significant, tested at any conventional levels.

Regarding the sign of the treatment variable, our finding in height outcome is in line with Mangyo (2008). However, our estimates in the BMI model with the presence of the endogeneity problem are contradictory. Mangyo (2008) indicated that having on-premises piped water negatively affects child’s BMI in China, but our study found its positive impact. Jalan and Ravallion (2003) indicated that children with access to public tap sources might face a higher duration of diarrhea than their friends who have on-plot piped water. They also highlighted that access to piped water in the yard could benefit households by lowering the possibility of contamination owing to storage, hence reducing the duration of adverse illness. And then, the health outcomes of children could be improved over time. As noted previously, however, the estimate of having a piped water supply in the home is not statistically significant when using the proposed information to determine children's height and BMI.

4.2. The role of mother’s education on health gains from access to on-plot piped water

We continue to examine the role of mothers in the relationship between in-yard piped water accessibility and health outcomes measured by BMI and height. The proposed estimators will be separately used for low-educated and educated mothers. Poorly educated mothers are classified as those who have never attended school in all rounds. Notably, Vietnam has the highest rate of educated mothers at 90.25 percent, followed by Ethiopia at 57.92 percent and India at 52.8 percent. Table 4 displays the estimated results using the FE and FEIV approaches.

In the first-stage regression using the FEIV method, we found that the coefficients of the leave-out ratio in piped water sources of the educated-sample models are higher than that of the less-educated-sample models. This reflects that mothers with higher levels of education will have much greater access to piped water sources than those with lower levels of education. Mangyo (2008) explained that more decisive political influence makes it simpler for women with higher qualifications to access water projects funded by the government or NGOs. Women with more education often live close to social and political locations, which lets them benefit from water projects sooner than women with less education (Mangyo, 2008).

The F statistics on the excluded instruments are well above 10, indicating that the selected IV is strong. Besides, the statistics from the Hausman test are insignificant, tested at any significance level. This suggests that the treatment variable, piped water accessibility of a household, is exogenous when modelling these health indicators. This supports the use of a FE approach.

Using a FE estimator, we find the positive and statistically significant coefficient of piped water accessibility in the height model with the educated-mothers sample. Among children from educated mothers, a child with piped water supply in the home will be expected to have 0.034 higher in height than his/her peers, on average. However, the less educated mothers sample shows a negative and insignificant estimate in the height model. Regarding the BMI regressions, both educated mother’s and less educated mother’s samples show that having piped water supply in the home will positively support the BMI indicator of children. Moreover, the findings in BMI regressions using a FE estimator are insignificant. Moreover, in absolute terms, the educated sample has a greater magnitude impact of on-premises piped water accessibility on the BMI outcome than the less educated group. This partly reflects the finding that higher inequality in piped water access may reduce its BMI-health impact in the sample of highly educated women.

Our findings still support the argument of Mangyo (2008) that mother education and piped water accessibility are complementary in producing child health. The maternal educational level reflects their hygienic water handling. Mothers with higher educational levels could adopt skills and knowledge in water storing or boiling better and more efficiently, subsequently positively affecting child health (Lamichhane and Mangyo, 2011). Consequently, anthropometric outcomes are conditional on mothers’ education, especially for poor households (Jalan and Ravallion, 2003). Similar to Jalan and Ravallion's (2003) and Mangyo's (2008) studies, we are unable to assess quantitatively how maternal education turns into actual water treatment and water quality and, consequently, how that influences child health due to the data unavailability.

5. Robustness check

5.1. Testing for instrumental variables

As discussed above, one of the most critical assumptions of instruments is the exclusive restrictions, meaning that there is no correlation between the selected instrumental variable and the error terms (ב_it). Theoretically, the exclusive restriction is violated in our study if the community-level access to clean water is subject to endogenous project placement (Mangyo, 2008 and Lamichhane and Mangyo, 2011). Consequently, the estimates will be inconsistent if our selected instrumental variable is endogenous. To solve this issue, we investigate the water project placement using the data on the disease symptoms indicated by Mangyo (2008). Due to data availability, we cannot construct panel data on disease symptoms instead of providing cross-sectional data for Round 2 in both groups. Four disease symptoms, including malaria, pneumonia, diarrhea, and asthma, will be divided into the water-relevant group (malaria and diarrhea) and the water-irrelevant group (pneumonia and asthma).

The data for India show the highest percentage of children facing water-relevant symptoms with around 14.1%; meanwhile, the highest figure for the percentage of children facing water-irrelevant symptoms is for Ethiopia with 2%. Comparing the two groups of low- and high-educated mothers, we find that the difference in these variables between the two groups in each country is not statistically significant (Table S4, see Supplementary materials file). Similar to the study of Lamichhane and Mangyo (2011), we use the survey probit instrumental regression approach to see the impact of the treatment variable (water accessibility) on the probability of facing disease symptoms. The estimated results are presented in Table 5.

Accordingly, the marginal effects of the treatment variable on the incidence of water-irrelevant and water-relevant disease symptoms are not statistically significant. Besides, our findings indicate a negative relationship between the treatment variable and the incidence of water-relevant disease symptoms in both educated and less-educated samples, but not statistically significant. Our findings are different from the results of Mangyo (2008) and Lamichhane and Mangyo (2011) when they found positive and statistically significant estimates in the water-relevant regressions using the less-educated sample. Our finding using the data on the disease symptoms suggests that water projects could be placed exogenously across the surveyed communities, leading to the validity of the selected instrumental variable, the leave-out ratio in piped-water supply.

The main result section indicates that a FE estimator is appropriate with the result of the Hausman test that we cannot reject the null hypothesis of exogenous treatment variable. Therefore, we should not use the selected instrumental variable, the leave-out ratio in piped water supply. We continue to test the robustness check by using another instrumental variable which is the leave-out ratio in the most popular source. The most popular water sources are bore well or protected well sources in Vietnam and public standpipe sources in Ethiopia and India. These water supply sources are mainly used in these selected countries but tend to decrease over time (see Table 1). First, we run the regression with solely new instrumental variable, reported in FEIV2 models (see Table 6). And then, we jointly use both leave-out strategy indicators as instruments of the endogenous variable in FEIV3 models.

The first stage results in FEIV2 and FEIV3 models show the negative and statistically significant impacts of the leave-out ratio in the most popular source on the treatment variable.

Moreover, using solely the second instrumental variable, the estimates of the treatment variable tend to be bigger in modelling health outcomes compared to the FEIV regressions reported in columns 1 and 4 in Table 6. When we use jointly both instrumental variables, the estimates tend to be unchanged. Moreover, the Sargan-Hansen test for over-identification in both models using jointly two IVs indicates that the proposed IVs are exogenous, tested at the 5% level of significance. Regarding the Hausman test, the FEIV3 regressions in both height and BMI outcomes show that the treatment variable could be exogenous, tested at a 5% level of significance. The results still suggest that a FE estimator is appropriate to estimate the causal effect of having piped water supply on health outcomes of children. The results are consistent with the conclusion about not using the instrumental variable in the main section because of the exogenous treatment variable.

5.2. The exclusion or inclusion some time-variant regressors

Following the study of Dell *et al.* (2014), we continue to conduct another sensitive test by excluding or including some important variables which are time-variant. First, we exclude the list of regressors reflecting the community's characteristics. They reflect the variation of water pollution in each community, the local population size, the ability to face epidemics that affect humans, and access to factories. Obviously, these variables are potentially correlated with both the selected instrument and the health outcomes of children. The results are reported in Specification 1 in Table 7 using both FE and FEIV approaches.

Accordingly, the results of the Hausman test suggest that a FE estimator is preferred to estimate the impact of water accessibility on child height, while the BMI regression should use the FEIV. Particularly, the estimates of the treatment variable change a bit. In the height model with the FE estimator, the estimate of piped water accessibility drops from 0.014 in the regression using community variables to 0.012 in the new regression without such these variables. In the BMI model, the estimate of piped water accessibility in the regression excluding regressors reflecting community's characteristics is overestimated with 0.029, higher than 0.019 in the base model with the FE method. The finding suggests that removing the effect of time-invariant unobserved community variables, the estimates could still be biased due to the possible correlation between time-variant unobserved variables at the community level with child health and water accessibility. Indeed, a reverse change in the level of local water pollution from industrial waste, local family garbage, pesticides, and fertilizers in local agricultural lands could reduce child health and increase the possibility of installing water access, creating a negative bias. Therefore, our proposed method with the time-variant community characteristics remains robust.

We continue to include the number of time children spend on their sleeping and their leisure activities into the main regressions. Following Jacobson (2000), the child health production function will be respected to the time restriction. It means that there are potential relationships between how many hours a child spends on his/her sleeping and leisure activities and his/her health outcomes. Besides, if a family installs the piped water supply inside their home, the fetching water time for mothers could be significantly reduced, affecting the time children spend on their own activities. The results are reported in Table 7. Accordingly, specification 2 will include only the number of hours per day a child spends sleeping (in the log form) in the main regressions. In specification 3, instead of using the sleeping time of children, we use the number of hours per day a child spends on leisure activities. And then, the specification 4 will jointly using both variables.

The Hausman tests at the bottom of Table 8 in these specifications indicate that the treatment variable is exogenous in modelling health outcomes, leading to the use of a FE estimator. The estimates of piped water accessibility in all regressions are positive and insignificant. Besides, we also find that the impact's magnitude of water accessibility on health in adjusted models plausibly varies little compared to the models in the main results. Therefore, our main regressions remain robust.

5.3. Using an alternative measure of health outcomes: binary variables

We replace the measure of anthropometric outcomes with binary variables. Notably, a YL child with a height-for-age Z-score below -2 (relative to international medians for well-nourished populations) is considered stunted. Besides, a thin child is with a BMI-for-age Z-score below -2. We continue to use the FE and FEIV approaches for the binary outcomes. Alternatively, we use the bivariate probit regression (BPIV) and the probit regression for analysis. The BPIV method captures the nonlinear effect of water accessibility, so its estimate is reasonable. Besides, the method also allows the marginal effect of the treatment variable on the probability of stunting and thinness to be dependent on other observed regressors while the FEIV method with the linear assumption does not allow that. Because of the non-linear models, the Average partial effect (APT) of the treatment variable in BPIV models will be obtained from bootstrap samplings. Table 8 reports the impacts of on-plot piped water on stunted and thin status of children for the whole sample, using both FEIV and BPIV methods. We also report the results using a FE estimator and the probit method for a panel data. For examining the second question of the role of maternal education, the results are reported in Table 9.

In Table 8, the Hausman test using the FEIV estimator cannot reject the null hypothesis that water accessibility is exogenous. We should not use the selected instrumental variable. Therefore, we preferred using a FE estimator. As discussed above, because the outcomes are binary outcomes, the probit estimator is appropriated. Accordingly, we find a negative and statistically significant coefficient of the treatment variable in the stunted regression. It means that the probability of a child facing stunted will be 17% lower if he/she has piped water supply in the home. Meanwhile, the BMI regression provides an insignificant estimate.

In Table 9, we find that all models' estimated coefficients of piped water accessibility are insignificant using the FE method. The probit method shows a negative and statistically significant estimate in the stunted model with educated-mothers sample. Meanwhile, other regressions display insignificant estimated coefficients of piped water accessibility using the probit estimator. The results also support our previous finding that the health effect of access to piped water supply primarily targeted children of highly qualified mothers.

5.4. Attrition bias

In this section, we discuss sample attrition bias. For analysis, we drop households having relocated activities, which equates to around 17% of the observations from the initial data, to keep a sample of children who stayed in the same community. As a result, we can face attrition bias, leading to inaccurate estimates. The migration in developing countries could explain the reduction in the sample size. People, especially the poor, probably try to move out of their hometowns to find more opportunities in urban areas or big cities.

To check if sample attrition is systematically in our study, we regress the treatment variable on the health outcomes by using two samples: with and without YL children who moved out of their hometowns. The FE estimator is used instead of FEIV because instrumental might not be measured in the sample including migrant YL children. In this case, we cannot calculate the leave-out ratio without the participation of other children in the new community. Therefore, we do not discuss the endogeneity problem of the treatment variable in this section. The results are presented in Table S5 (see *Supplementary materials file*). Accordingly, after adding a sample of migrant YL children, we find the same sign of the coefficients of the treatment variable compared to the sample without such children. Besides, the impact magnitudes of the estimated treatment variable from the two samples mentioned above are not too different in each model. Therefore, we partially conclude that attrition bias might not be a big problem in our study.

6. Heterogeneity

Unsafe drinking water is considered a primary cause of diarrhea, resulting in millions of deaths of roundly-five-year children (Jalan and Ravallion, 2003). The health of pre-schoolers largely depends on the water quality. Therefore, piped water supply could be the best solution to the death-due-to-diarrhea problem in this group. To investigate the heterogeneous effect across ages, we separately estimate the impact of on-premises piped water access on anthropometric outcomes by age groups of roughly five years old, eight years old, 12 years old, and 15 years old. The results are presented in Tables S6 and S7 (see *Supplementary materials file*).

Accordingly, the Hausman test suggests that water accessibility is endogenous in most of the regressions, only excluding the BMI regression with a 15-years-old sample. Moreover, the F tests are greater than 10, leaving a strong IV. The FEIV estimates of water accessibility on height are positive and statistically significant at 5% level across all the age groups. Modelling the BMI outcome, we also found the positive and statistically significant estimates of water accessibility in all age samples, excluding for the 15-years-old samples. Notably, piped water accessibility affects anthropometric measures of children in different age groups very differently. Particularly, the impact's magnitude of the treatment variable in the pre-schoolers group relatively higher than that of other groups. Around-5-years-old children living with on-site piped water supply might have a significantly higher height and BMI of 0.599 units and 0.869 units, respectively, than those without such water sources. Meanwhile, the figures for children who are around 8 years old are 0.393 and 0.564 units in the height and BMI regressions, respectively. This may explain that the level of dependence on clean water of children at around-5-years-old age period is higher than that of older children.

We then examine the heterogenous impact of piped water accessibility on health by gender. Table S8 (see *Supplementary materials file*) provides the estimated results of whether having an in-yard piped water source affects health outcomes in boys and girls separately by using FE or FEIV method. The results of the Hausman test indicate that piped water accessibility could be exogenous in all regressions; therefore, we will report results from the FE estimator. Particularly, we do not find any significant impacts of piped water accessibility on health of boys and girls. However, we can see that the boy sample shows higher impact's magnitude of the treatment variable on height than a group of girls (in the absolute value). Meanwhile, the BMI regression using the FE method show a higher impact's magnitude (in the absolute value) in girls than in boys. Biologically, the level of brawn in boys is more significant than in girls.

Furthermore, boys will gain more brawn than girls if the water source in their yard is piped water (Zhang and Xu, 2016).

7. Conclusion

The study examined the impact of in-yard piped water accessibility on child health as measured by anthropometric indicators with the endogeneity problem in developing countries. We use YLS panel data to analyse thousands of children in Ethiopia, India, and Vietnam. The data is suitable for estimating the impact of change in piped water accessibility on health outcomes, primarily focusing on the poor groups. The endogeneity problem of water accessibility is a concern of this study, leading to the application of using instruments. Following previous studies by Mangyo (2008) and Lamichhane and Mangyo (2011), we use the leave-out strategy in piped water sources as a main instrumental variable. Water projects could be randomly allocated across surveyed poor communities. However, the Hausman test of the FEIV estimator supports that the treatment variable is exogenous, leading to not using the instrumental variables. This is due to the use of data that is mainly reflective of poor and disadvantaged communities and not the entire population. Access to water may be exogenous to the personal purpose of households because water projects are mainly undertaken by local and central governments or non-governmental organizations. (Mangyo, 2008). The findings are based on the FE specification.

The study finds that the percentage of households having in-yard piped water supply tends to increase over time across countries but remain low. Besides, the FE estimator indicates that access to piped water sources inside the dwelling supports increasing height and BMI indicators of children in developing countries. Nevertheless, our findings are not statistically significant. Regarding the role of maternal education, children from educated mothers have more chances to access piped water sources than their peers from lower-qualification mothers. The health effect of piped water accessibility also primarily targeted children of highly qualified mothers in these countries in terms of height. Our findings support the argument of Mangyo (2008) about the complementary relationship between mother education and water accessibility. Having piped water inside the dwelling might support reducing the risk of facing diarrhea, subsequently enhancing child health.

In developing countries, namely Ethiopia, India, and Vietnam, their safe water coverages are close to reaching the whole country. However, their current policies and programmes have not specified the role of in-yard piped water sources, especially for the poor. The higher coverage mainly depends on outside safe water sources, which could be more contaminated during storage and handling than in-dwelling piped water, negatively affecting child health. Therefore, our findings motivate policymakers to pay more attention to expanding in-yard piped water for the whole population. For the poor, we need more specific support to access this water source. We also agree with Mangyo’s argument that improving mothers’ health knowledge regarding safe drinking water may be a crucial component of policies aiming to promote the positive impacts of safe water and sanitation in developing nations.

Our study might have some potential limitations. First, our data is not representative of the population of selected developing countries. Therefore, our work cannot fully answer the question of how having piped water supply in the home affects child development in developing countries. Second, we cannot control the quality of piped water due to data unavailability. Ethiopia, India, and Vietnam are listed among the top countries with low-quality drinking water, according to the report of Yale University about the Environmental Performance Index (Yale, 2022), leading to adverse effects on the health and development of children. We also cannot examine qualitatively how maternal education translates into actual water treatment and water quality and, subsequently, how this impacts child health. Looking at the causal effect of water quality on health might support developing countries in developing and implementing more effective policies.

8. Data availability statement

The study uses the data for analysis from the Young Lives project, aiming to exploit the changing nature of childhood poverty in four countries: Ethiopia, India, Peru, and Vietnam. The project is funded by The UK's Department for International Development (DFID) and Irish Aid. All information and raw data of the project can be found at www.younglives.org.uk upon reasonable request.

9. Conflict of interest

None declared

10. Acknowledgements

We are grateful to Dr. Jonas Fooken, Dr. Jean Spinks, Dr. Danusha Jayawardana, Dr. Sabrina Lenzen, Dr. Lan Nguyen, Michelle Tran (ANU), Aarushi Dhingra for their helpful and invaluable comments at the seminar at the Centre for the Business and Economics of Health, The University of Queensland. We also thank Professor. Denzil Fiebig (The University of New South Wales) and all audiences and participants at the 43rd Annual Australian Health Economics Society (AHES) in Brisbane, Australia for their helpful comments. We acknowledge the support from The University of Queensland. We thank the UK Data Service for providing the data that we used in this study.

11. Credit author statement

The paper is a chapter of the Ph.D. thesis of Dao Nguyen Dinh titled *"The economics of child health: Determinants of child physical and mental health and their long-term consequences"* at the University of Queensland, Australia. Dao Nguyen Dinh developed the methodology, acquired the data and performed all data analysis, in addition to conducting the literature review, and drafting of the original draft. Luke Connelly, Stephen Birch, and Ha Nguyen advised on the aims of the project, methodology and interpretation, investigated data analysis, edited the revised drafts of the paper, and provided full supervision to Dao Nguyen Dinh.

Reference

ADB (2010), *Vietnam: Water and sanitation sector assessment, strategy, and roadmap.* Southeast Asia Department Working Paper. Asian Development Bank.

Ali, M. (2019). Multisectoral approach to nutrition in Pakistan. *Journal of Water Sanitation and Hygiene for Development*, *9*(1), 152–163.

Allen, E., Maria Morazan, I., & Witt, E. (2018). Actively engaging women is helping solve the global water crisis. *Journal of Water Sanitation and Hygiene for Development*, *8*(4), 632–639.

Becker, G.S., 1965. A theory of the allocation of time. Economic Journal 75, 493–517

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., & Rivera, J. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet*, *371*(9608), 243–260.

Cairncross, S., Hunt, C., Boisson, S., Bostoen, K., Curtis, V., Fung, I. C. H., & Schmidt, W. P. (2010). Water, sanitation and hygiene for the prevention of diarrhoea. *International Journal of Epidemiology*, *39*(SUPPL. 1).

Conant, J., & Fadem, P. 2008. *A community guide to environmental health.* The Hesperian Foundation, USA.

Devoto, F., Duflo, E., Dupas, P., Parienté, W., Pons, V. (2012). Happiness on tap: piped water adoption in urban Morocco. *American Economic Journal: Economic Policy* 4, 68-99.

Dewey, K. G., & Begum, K. (2011). Long-term consequences of stunting in early life. *Maternal and Child Nutrition*, *7*(SUPPL. 3), 5–18.

Grossman, M., (1972). On the concept of health capital and the demand for health. *Journal of Political Economy 82*(2), 223–255.

Headey, D., & Palloni, G. (2019). Water, Sanitation, and Child Health: Evidence From Subnational Panel Data in 59 Countries. *Demography*, *56*(2), 729–752.

Jacobson, L. (2000). The family as producer of health - An extended grossman model. *Journal of Health Economics*, *19*(5), 611–637.

Jalan, J., & Ravallion, M. (2001). Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics, 112*(1), pp. 153-173.

Lamichhane, D. K., & Mangyo, E. (2011). Water accessibility and child health: Use of the leave-out strategy of instruments. *Journal of Health Economics*, *30*(5), 1000–1010.

Le, H. T., & Nguyen, H. T. (2018). The impact of maternal mental health shocks on child health: Estimates from fixed-effects instrumental variables models for two cohorts of Australian children. *American Journal of Health Economics*, *4*(2), 185–225.

Leszczensky, L., & Wolbring, T. (2019). How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. *Sociological Methods and Research*.

Mangyo, E. (2008). The effect of water accessibility on child health in China. *Journal of Health Economics*, *27*(5), 1343–1356.

Mills, J. E., Cumming, O. (2016), *The impact of water, sanitation and hygiene on key health and social oitcomes: Review of evidence*, UNICEF

Overbo, A., Williams, A. R., Evans, B., Hunter, P. R, & Bartram, J., (2016). On-plot drinking water supplies and health: A systematic review, *International Journal of Hygiene and Environmental Health*, *219,* 317-330.

Ray, I. (2007). Women, water, and development. *Annual Review of Environment and Resources, 32*, 421–449.

Ricciuti H.N. (1981) “Developmental consequences of malnutrition in early childhood”, *The Uncommon Child. Genesis of Behavior*, edited by R.J. Sternberg & L. Okagaki,vol 3. Springer, Boston, MA.

Schott, W., Aurino, E., Penny, M. E., & Behrman, J. R. (2019). The double burden of malnutrition among youth: Trajectories and inequalities in four emerging economies. *Economics and Human Biology*, *34,* 80–91.

Sharma, A. (2017). Drinking Water Quality in Indian Water Policies, Laws, and Courtrooms: Understanding the Intersections of Science and Law in Developing Countries. *Bulletin of Science, Technology and Society, 37(1)*, 45-56.

Stock, J. H., and Yogo, M. (2005). “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models*, edited by Donald W. K. Andrews, 80–108. New York: Cambridge University Press

UNICEF (1990). Strategy for improved nutrition of children and women in developing countries. *UNICEF Policy Review*, *1*(November 1989), 13–24.

UNICEF (2018), *One wash national program (OWNP): A multi-sectoral SWAP, Review of Phase 1.* The United Nations International Children’s Emergency Fund.

UNICEF (2020), *Policy brief: Water, sanitation and hygiene in Vietnam.* The United Nations International Children’s Emergency Fund

van Cooten, M. H., Bilal, S. M., Gebremedhin, S., & Spigt, M. (2019). The association between acute malnutrition and water, sanitation, and hygiene among children aged 6–59 months in rural Ethiopia. *Maternal and Child Nutrition*, *15*(1), 1–8.

Wells, J. C., Sawaya, A. L., Wibaek, R., Mwangome, M., Poullas, M. S., Yajnik, C. S., & Demaio, A. (2020). The double burden of malnutrition: aetiological pathways and consequences for health. *The Lancet*, *395*(10217), 75–88.

WHO (2009). *AnthroPlus for Personal Computers. Manual: Software for assessing growth of the world's children and adolescents*, World Health Organization, http://www.who.int/growthref/tools/en/

WHO (2014). *Preventing diarrhoea through better water, sanitation and hygiene*. World Health Organization.

WHO (2019), *World health statistics 2019: monitoring heakth for the SDGs*. World Health Organization.

WHO (2020), *World health statistics 2020: monitoring heakth for the SDGs*. World Health Organization

Wooldridge, J. M. (2012). *Introductory econometrics : a modern approach* (5th ed.). South-Western Cengage Learning

World Bank (2014), *Water supply and sanitation in Vietnam: Turning finance into services for the future.* World Bank.

Yale (2022). *Sanitation and Drinking Water: Environmental Performance Index*. Yale Center for Environmental Law and Policy, Yale University. https://epi.yale.edu/epi-results/2020/component/h2o

Zhang, J. & Xu, L. C. (2016). The long-run effects of treated water on education: The rural drinking water program in China. *Journal of Development Economics, 122,* 1-15.

Appendix

Table 1: Descriptive statistics

	Young Cohort				Old Cohort	
	Round 2	Round 3	Round 4	Round 5	Round 2	Round 3
VIETNAM						
Piped water accessibility	0.09 (0.28)	0.13 (0.34)	0.20 (0.40)	0.27 (0.44)	0.12 (0.32)	0.18 (0.39)
The leave-out ratio in piped water source	0.09 (0.20)	0.13 (0.24)	0.20 (0.32)	0.27 (0.34)	0.12 (0.24)	0.18 (0.28)
The leave-out ratio in the most popular source	0.53 (0.38)	0.48 (0.39)	0.50 (0.41)	0.44 (0.39)	0.48 (0.37)	0.44 (0.37)
INDIA						
Piped water accessibility	0.10 (0.30)	0.16 (0.36)	0.23 (0.42)	0.24 (0.42)	0.11 (0.31)	0.19 (0.39)
The leave-out ratio in piped water source	0.10 (0.17)	0.16 (0.22)	0.23 (0.24)	0.24 (0.24)	0.11 (0.18)	0.19 (0.27)
The leave-out ratio in the most popular source	0.56 (0.32)	0.51 (0.35)	0.31 (0.24)	0.27 (0.21)	0.54 (0.34)	0.50 (0.34)
ETHIOPIA						
Piped water accessibility	0.09 (0.28)	0.13 (0.34)	0.20 (0.40)	0.25 (0.43)	0.13 (0.33)	0.18 (0.39)
The leave-out ratio in piped water source	0.09 (0.15)	0.13 (0.19)	0.20 (0.24)	0.25 (0.26)	0.13 (0.18)	0.18 (0.22)
The leave-out ratio in the most popular source	0.37 (0.30)	0.34 (0.31)	0.22 (0.27)	0.28 (0.31)	0.41 (0.30)	0.36 (0.29)

Note: Standard errors are shown in parentheses

Table 2: Summary statistics by piped water accessibility

	Without piped water (1)	With piped water (2)	Diff. (1)-(2)
Height-for-age z-score	-1.405	-1.019	-0.386***
Height-for-age z-score	-1.132	-0.894	-0.239***
Thinness	0.230	0.205	0.025***
Stunting	0.290	0.174	0.117***
Child gender (Male)	0.521	0.531	-0.010
Urban areas	0.150	0.660	-0.510***
Child to household size ratio	0.462	0.415	0.046***
The percentage of women in each household	0.262	0.284	-0.022***
Mom age	35.623	37.648	-2.025***
Food expenditure per capita (log form)	2.031	2.239	-0.208***
Nonfood expenditure per capita	1.566	2.226	-0.660***
Consumer durable index	0.348	0.516	-0.167***
Whether a YL child faces the shock-illness of a father or mother	0.296	0.302	-0.006
Have a household faced any natural disasters since the previous round?	0.265	0.087	0.178***
Population size in the commune (log form)	8.562	8.990	-0.428***
Problem: Is local families' garbage dumped at these water sources?	0.337	0.386	-0.049***
Problem: uses of pesticides and fertilizers in local agricultural lands and faces the problem of standing water, open drains?	0.194	0.129	0.066***
Service availability: Children's playground	0.473	0.507	-0.033***
Whether the locality have any factories that employ their resident?	0.528	0.611	-0.083***

* p<0.1; ** p<0.05; *** p<0.01

Table 3: On-premises piped water accessibility and child height and BMI

	Height-for-age z-score			BMI-for-age z-score		
	OLS (1)	FE (2)	FEIV (3)	OLS (4)	FE (5)	FEIV (6)
Piped water accessibility	0.082*** (0.025)	0.014 (0.016)	0.049 (0.054)	0.032 (0.030)	0.019 (0.024)	0.148* (0.084)
Time dummies	No	Yes	Yes	No	Yes	Yes
Country dummies	No	Yes	Yes	No	Yes	Yes
Covariates						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family's characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Community's characteristics	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.10	0.04	0.04	0.11	0.12	0.11
Observations	21,387	21,387	21,303	21,336	21,336	21,247
The first stage: Dependent variable is piped water accessibility						
The leave-out ratio in piped water source			0.741*** (0.028)			0.738*** (0.028)
Cragg-Donald Wald F statistic			1437.760			1419.799
Kleibergen-Paap rk Wald F statistic			687.058			685.321
p-value of F test of excluded instruments			0.000			0.000
Hausman test (p value)			0.494			0.108

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage of the FEIV approach is pipedw, whether household access to piped water into dwelling/yard/plot? An instrumental variable used in columns (3) and (6) is the leave-out commune ratio in piped water sources on the premises. Control variables reflect the characteristic of children, their families, and their communities. In the OLS model, we also include time-invariant variables such as child gender and whether a household lives in urban areas.

Table 4: On-premises piped water accessibility and child anthropometric outcomes: The role of education

	Height-for-age z-score				BMI-for-age z-score			
	Educated		Less educated		Educated		Less educated	
	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE
Piped water accessibility	0.050	0.034*	0.025	-0.048	0.132	0.008	-0.115	0.004
	(0.059)	(0.019)	(0.132)	(0.030)	(0.093)	(0.028)	(0.199)	(0.045)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.04		0.04		0.08		0.21	
Observations	14,559	14,602	6,670	6,710	14,524	14,567	6,653	6,698
The leave-out ratio in piped water source	0.781***		0.599***		0.777***		0.601***	
	(0.033)		(0.054)		(0.033)		(0.053)	
Cragg-Donald Wald F statistic	1110.089		271.923		1086.465		277.348	
Kleibergen-Paap rk Wald F statistic	561.680		120.971		549.839		127.891	
p value of F test of excluded IVs	0.000		0.000		0.000		0.000	
Hausman test (p value)	0.779		0.564		0.162		0.536	

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage of the FEIV approach is pipedw, whether household access to piped water into the dwelling/yard/plot? Control variables reflect the characteristic of children, their families, and their communities. We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported because of the exactly identified regression model.

Table 5: On-premises piped water accessibility and disease symptoms

	Water-relevant disease symptoms		Water-irrelevant disease symptoms	
	Educated	Less educated	Educated	Less educated
Piped water accessibility	-0.269 (0.409)	-0.426 (0.890)	0.043 (0.442)	1.746 (1.061)
Marginal effect (treatment variable)	-0.025 (0.038)	-0.068 (0.086)	0.001 (0.015)	0.083 (0.076)
Time dummies	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes
Covariates				
Child characteristics	Yes	Yes	Yes	Yes
Family's characteristics	Yes	Yes	Yes	Yes
Community's characteristics	Yes	Yes	Yes	Yes
Observations	4,587	2,268	4,587	2,268
The first stage: Dependent variable is piped water accessibility				
The leave-out ratio in piped water source	0.816*** (0.039)	0.496*** (0.079)	0.816*** (0.039)	0.496*** (0.079)

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage is pipedw, whether household access to piped water into the dwelling/yard/plot? Control variables reflect the characteristic of children, their families, and their communities. We solely run the survey probit instrumental regression approach (by using svy:ivprobit) with an IV which is the leave-out commune ratio in piped water sources on the premises.

Table 6: Robustness check: Using second instrumental variable

	Height-for-age z-score			BMI-for-age z-score		
	FEIV (1)	FEIV2 (2)	FEIV3 (3)	FEIV (4)	FEIV2 (5)	FEIV3 (6)
Piped water accessibility	0.049 (0.054)	0.063 (0.139)	0.050 (0.054)	0.148* (0.084)	0.471** (0.203)	0.155* (0.084)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
Covariates						
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Family’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Community’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.04	0.04	0.04	0.11	0.09	0.11
Observations	21,303	21,303	21,303	21,247	21,247	21,247
The first stage: Dependent variable is piped water accessibility						
The leave-out ratio in piped water source	0.741*** (0.028)		0.726*** (0.030)	0.738*** (0.028)		0.724*** (0.030)
The leave-out ratio in the most popular source		-0.175*** (0.012)	-0.023** (0.011)		-0.174*** (0.012)	-0.022* (0.011)
Cragg-Donald Wald F statistic	1437.760	223.988	721.880	1419.799	217.920	711.597
Kleibergen-Paap rk Wald F statistic	687.058	217.603	359.935	685.321	210.507	358.218
p-value of F test of excluded instruments	0.000	0.000	0.000	0.000	0.000	0.000
Hansen J test (p value)			0.915			0.087
Hausman test (p value)	0.494	0.724	0.493	0.108	0.024	0.085

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage of the FEIV approach is pipedw, whether household access to piped water into dwelling/yard/plot? An instrumental variable used in columns (1) and (4) is the leave-out commune ratio in piped water sources on the premises, while regressions (2) and (5) use the leave-out ratio in the most popular source. The FEIV results using jointly both selected Ivs are reported in columns (3) and (6). Control variables reflect the characteristic of children, their families, and their communities

Table 7: Robustness check: Excluding or including some important variables

	Specification 1				Specification 2				Specification 3				Specification 4			
	Height		BMI		Height		BMI		Height		BMI		Height		BMI	
	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE
Piped water accessibility	0.025	0.012	0.229***	0.029	0.048	0.014	0.150*	0.019	0.050	0.014	0.140*	0.019	0.050	0.013	0.142*	0.019
	(0.054)	(0.016)	(0.085)	(0.023)	(0.054)	(0.016)	(0.084)	(0.024)	(0.054)	(0.016)	(0.084)	(0.024)	(0.054)	(0.016)	(0.084)	(0.024)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates																
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community's characteristics	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sleeping time	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Leisure activities time	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.04	0.04	0.10	0.11	0.04	0.04	0.11	0.12	0.04	0.04	0.12	0.12	0.04	0.04	0.12	0.12
Observations	21,691	21,836	21,640	21,790	21,301	21,385	21,245	21,334	21,180	21,310	21,121	21,257	21,178	21,119	21,119	21,255
The first stage: Dependent variable is piped water accessibility																
The leave-out ratio in piped water source	0.741***		0.739***		0.740***		0.737***		0.739***		0.736***		0.739***		0.736***	
	(0.028)		(0.028)		(0.028)		(0.028)		(0.028)		(0.028)		(0.028)		(0.028)	
Cragg-Donald Wald F statistic	1488.698		1469.486		1435.292		1417.448		1424.671		1406.163		1422.387		1404.510	
Kleibergen-Paap rk Wald F statistic	711.464		709.829		685.491		683.827		678.832		677.393		677.720		676.312	
p-value of F test of excluded instruments	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	
Hausman test (p value)	0.797		0.012		0.503		0.105		0.472		0.131		0.478		0.126	

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage of the FEIV approach is pipedw, whether household access to piped water into the dwelling/yard/plot? Control variables reflect the characteristic of children, their families, and their communities. We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported because of the exactly identified regression model. Specification 1: Excluding the list of the regressors reflecting community's characteristics. Specification 2: Adding the number of hours per day a YL child spending in sleeping (in log form). Specification 3: Adding the number of hours per day a YL child spending on leisure activities. Specification 4: Adding both the numbers of sleeping and playing hours per day.

Table 8: On-premises piped water accessibility and binary outcomes: the whole sample

	Stunt				Thin			
	FEIV	FE	BPIV	Probit	FEIV	FE	BPIV	Probit
Piped water accessibility	-0.022	-0.007	-0.339***	-0.170***	0.010	-0.008	-0.179***	0.005
	(0.030)	(0.009)	(0.052)	(0.052)	(0.033)	(0.010)	(0.053)	(0.047)
APE			-0.096***				-0.046***	
			(0.015)				(0.013)	
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,303	21,387	21,386	21,387	21,247	21,247	21,335	21,336
The leave-out ratio in piped water source	0.741***		3.193***		0.738***		3.178***	
	(0.028)		(0.056)		(0.028)		(0.056)	
Cragg-Donald Wald F statistic	1437.760				1419.799			
Kleibergen-Paap rk Wald F statistic	687.058				685.321			
p value of F test of excluded IVs	0.000				0.000			
Hausman test (p value)	0.615				0.586			
rho			0.195***				0.110***	
			(0.033)				(0.034)	
Wald test (p_value)			0.000				0.000	

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Notes: Fully robust standard errors are shown in parentheses. The dependent variable in the first stage is pipedw, whether household access to piped water into dwelling/yard/plot? Control variables reflect the characteristic of children, their families, and their communities. We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported using the FEIV method because of the exactly identified regression model. Regarding the STATA command, we used biprobit for BPIV models. The Average partial effect (APE) of the treatment variable in BPIV models will be obtained from bootstrap samplings. Probit models: using xtprobit

Table 9: On-premises piped water accessibility and binary outcomes: the role of education

	Stunt				Thin			
	FE		Probit		FE		Probit	
	Educated	Less educated	Educated	Less educated	Educated	Less educated	Educated	Less educated
Piped water accessibility	-0.016 (0.010)	0.018 (0.019)	-0.235*** (0.062)	0.035 (0.094)	-0.006 (0.011)	-0.000 (0.021)	0.038 (0.056)	-0.061 (0.092)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,602	6,710	14,602	6,710	14,567	6,698	14,567	6,698

* p<0.1; ** p<0.05; *** p<0.01

Notes: Clustered standard errors are shown in parentheses. The dependent variable in the first stage is pipedw, whether household access to piped water into dwelling/yard/plot? Control variables reflect the characteristic of children, their families, and their communities. Probit models: using xtprobit

Appendix 2

Figure 1: Kdensity graph for Vietnam's samples

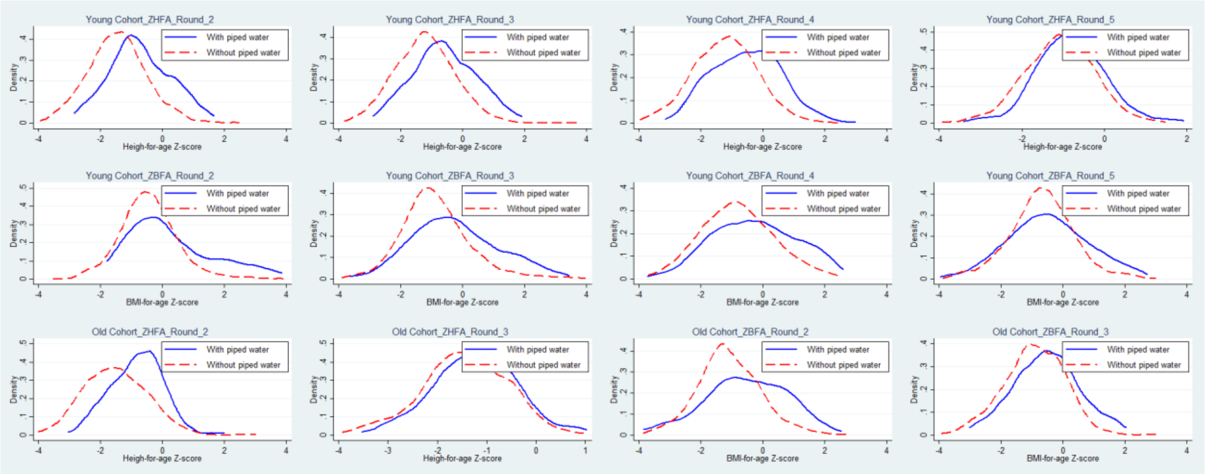


Figure 2: Kdensity graph for India's samples

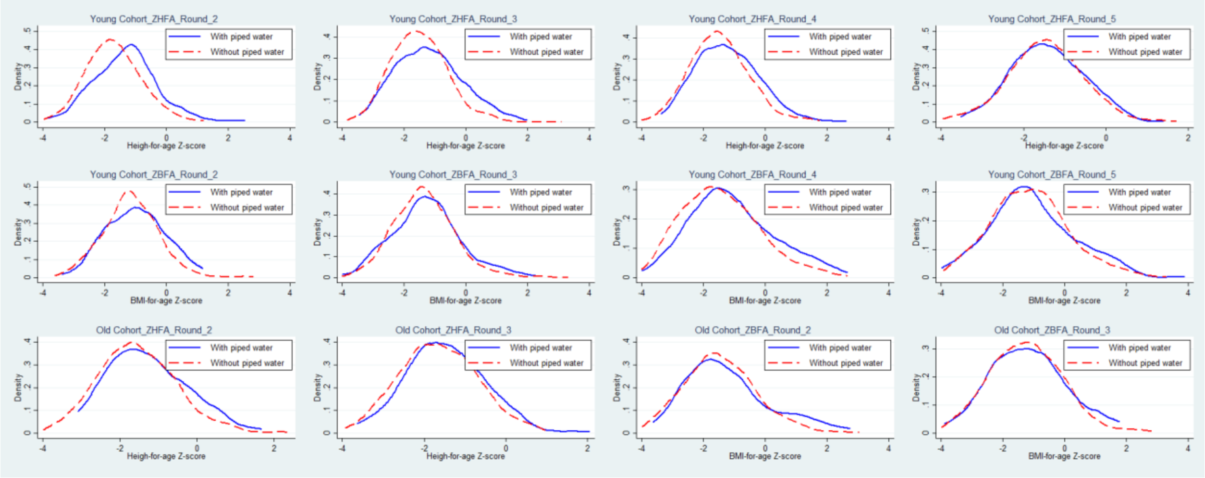
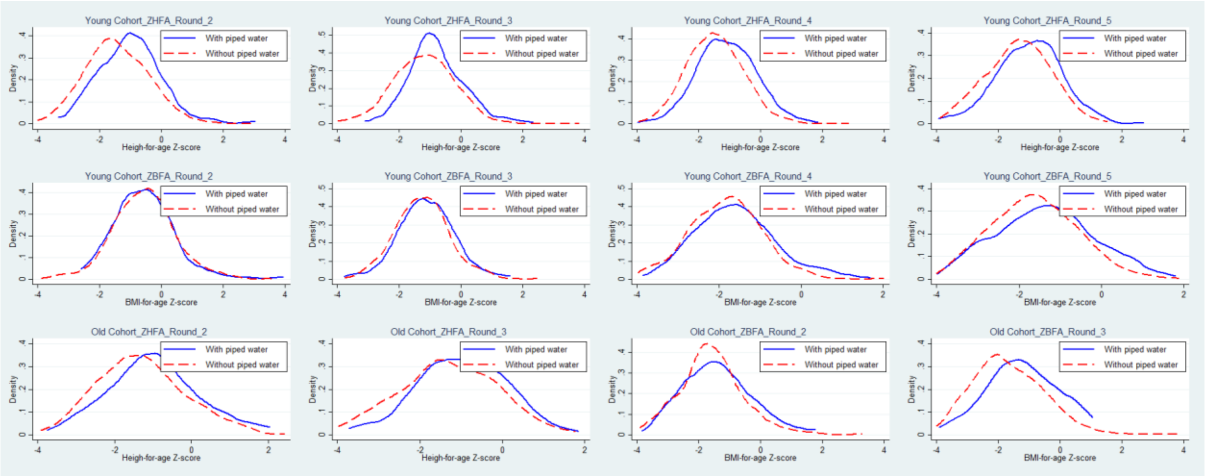


Figure 3: Kdensity graph for Ethiopia's samples



SUPPLEMENTARY MATERIALS

S1: OPTIMAL DEMAN FUNCTION OF HEALTH

The utility maximization problem: $\max\limits_{H,A,F,Z,I} U = \int_{t=0}^{T} e^{-\rho t} U(s_t, Z_t, TL_t)$ (A. 01)

subject to:

$$\dot{H}_t = I_t - \delta(t, X_t) * H_t \tag{A. 02}$$

$$\dot{A} = rA_t + Y(s_t) - \pi_t^I I_t - \pi_t^Z Z_t \tag{A. 03}$$

$$TW_t + TM_t + TZ_t + TL_t + s_t = 365 \tag{A. 04}$$

H_0 and A_0 are given; $H_T = H_{min};\ A_T \geq 0$

Hamiltonian function:

$$\mathcal{H} = e^{-\rho t} U(s_t, Z_t, TL_t) + \lambda_t [I_t - \delta(t, X_t) * H_t] + \mu_t [rA_t + Y(s_t) - \pi_t^I I_t - \pi_t^Z Z_t]$$

First-order conditions:

- ❖ $\frac{\partial \mathcal{H}}{\partial I_t} = 0 \rightarrow \lambda_t - \mu_t \pi_t^I = 0 \rightarrow \lambda_t = \mu_t \pi_t^I$ (A. 05)
  - $\rightarrow \dot{\lambda}_t = \dot{\mu}_t \pi_t^I + \mu_t \dot{\pi}_t^I$ (A. 06)
- ❖ $\frac{\partial \mathcal{H}}{\partial A_t} = -\dot{\mu}_t \rightarrow r\mu_t = -\dot{\mu}_t$ (A. 07)
  - $\rightarrow \mu_t = \mu_0 e^{-rt}$ (A. 08)
- ❖ $\frac{\partial \mathcal{H}}{\partial H_t} = -\dot{\lambda}_t \rightarrow e^{-\rho t} \frac{\partial U_t}{\partial s_t} \frac{\partial s_t}{\partial H_t} - \lambda_t \delta(t, X_t) + \mu_t \frac{\partial Y_t}{\partial s_t} \frac{\partial s_t}{\partial H_t} = -\dot{\lambda}_t$ (A. 09)

Because of $\frac{\partial U}{\partial h_t} \frac{\partial h}{\partial H_t} = 0$ from the pure investment model, A.10 becomes:

$$-\lambda_t \delta(t, X_t) + \mu_t \frac{\partial Y_t}{\partial s_t} \frac{\partial s_t}{\partial H_t} = -\dot{\lambda}_t \tag{A. 10}$$

(when $\frac{\partial U}{\partial h_t} \frac{\partial h}{\partial H_t} = 0$ from the pure investment model)

Substitute A.05; A.06; and A.08 into A.10; we have

$$\mu_t \frac{\partial Y_t}{\partial s_t} \frac{\partial s_t}{\partial H_t} = \mu_t \pi_t^I \delta(t, X_t) + r\mu_t \pi_t^I + \mu_t \dot{\pi}_t^I \tag{A. 11}$$

$$\rightarrow \quad \frac{\partial Y_t}{\partial s_t} \frac{\partial s_t}{\partial H_t} = \pi_t^I \delta(t, X_t) + r\pi_t^I + \dot{\pi}_t^I \tag{A. 12}$$

$s_t = \alpha_1 H_t^{-\alpha_2}$ so $\frac{\partial h_t}{\partial s_t} = -\alpha_1 \alpha_2 H_t^{-\alpha_2 - 1}$ (A. 13) and $\frac{\partial Y_t}{\partial s_t} = -w_t$ (A. 14)

Substitute (A.14); (A.15) into (A.13), we have:

$$\alpha_1 \alpha_2 w_t H_t^{-\alpha_2 - 1} = r\pi_t^I + \dot{\pi}_t^I + \pi_t^I \delta(t, X_t) \tag{A. 15}$$

Solving equation A.16, we will get the optimal demand function of H_t as follow:

$H_t^* = H(w_t, \pi_t^I, \delta(t, X_t), (\, r, \alpha_1, \alpha_2))$

Table S1: Summary statistics of selected variables

	Mean	Min	Max	Standard deviation		
				Overall	Between	Within
Height-for-age z-score	-1.34	-4	3.84	1.004	0.913	0.444
Height-for-age z-score	-1.092	-4	3.97	1.167	1.002	0.624
Thinness	0.225	0	1	0.418	0.333	0.271
Stunting	0.271	0	1	0.444	0.373	0.252
Child gender (Male)	0.522	0	1	0.499	0.5	0
Urban areas	0.236	0	1	0.424	0.427	0
Age of a YL child (in months, log form)	4.809	3.892	5.293	0.391	0.182	0.357
Child to household size ratio	0.454	0.077	1	0.148	0.121	0.091
The percentage of women in each household	0.266	0	1	0.11	0.089	0.071
Mom age	35.963	16	68	6.979	6.242	3.4
Food expenditure per capita (log form)	2.066	-1.001	8.936	0.705	0.58	0.392
Non-food expenditure per capita (log form)	1.677	-2.262	6.477	0.942	0.825	0.456
Consumer durable index	0.377	0	1	0.226	0.204	0.103
Whether a YL child faces the shock-illness of a father or mother	0.297	0	1	0.457	0.302	0.357
Have a household faced any natural disasters since the previous round?	0.235	0	1	0.424	0.303	0.31
Population size in the commune (log form)	8.634	2.996	11.156	0.925	0.883	0.294
Problem: Is local families' garbage dumped at these water sources?	0.345	0	1	0.475	0.338	0.346
Problem: uses of pesticides and fertilizers in local agricultural lands and faces the problem of standing water, open drains?	0.183	0	1	0.387	0.267	0.289
Service availability: Children's playground	0.479	0	1	0.5	0.359	0.357
Whether the locality have any factories that employ their resident?	0.542	0	1	0.498	0.382	0.322
Piped water accessibility	0.168	0	1	0.374	0.295	0.234
The leave-out ratio in piped water source	0.168	0	1	0.246	0.221	0.108
The leave-out ratio in the most popular water source	0.414	0	1	0.348	0.305	0.174

Table S2: The full second-stage results of Table 4 for the health impacts of piped water accessibility (Model FEIV)

	Height-for-age z-score			BMI-for-age z-score		
	Pool OLS (1)	FE (2)	FEIV (3)	Pool OLS (4)	FE (5)	FEIV (6)
Piped water accessibility	0.082***	0.014	0.049	0.032	0.019	0.148*
	(0.025)	(0.016)	(0.054)	(0.030)	(0.024)	(0.084)
Age of children in months (Log form)	-0.038	0.883***	0.885***	-0.627***	-2.482***	-2.477***
	(0.027)	(0.082)	(0.082)	(0.030)	(0.103)	(0.103)
Child to household size ratio	-0.183**	-0.096*	-0.097*	-0.066	0.061	0.057
	(0.073)	(0.053)	(0.053)	(0.081)	(0.073)	(0.073)
The percentage of women in each household	0.007	0.055	0.057	0.035	-0.024	-0.015
	(0.091)	(0.065)	(0.065)	(0.103)	(0.087)	(0.087)
Age of mother	0.025**	-0.044***	-0.044***	0.023**	0.271***	0.269***
	(0.011)	(0.012)	(0.012)	(0.012)	(0.017)	(0.017)
Age of mother square	-0.000**	-0.000***	-0.000***	-0.000	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Food expenditure per capita in real 2006 price, converted to USD (log value)	0.035**	0.044***	0.043***	0.156***	0.082***	0.080***
	(0.016)	(0.011)	(0.011)	(0.017)	(0.015)	(0.015)
Non-food expenditure per capita in real 2006 price, converted to USD (log value)	0.069***	0.027***	0.027***	0.048***	0.048***	0.047***
	(0.013)	(0.009)	(0.009)	(0.015)	(0.012)	(0.012)
Consumer durable index	0.611***	0.167***	0.167***	0.941***	0.134**	0.135**
	(0.054)	(0.046)	(0.046)	(0.062)	(0.062)	(0.062)
Have a household faced any natural disasters since the previous round?	-0.058***	0.012	0.011	-0.065***	0.024	0.021
	(0.018)	(0.012)	(0.012)	(0.019)	(0.016)	(0.016)
Whether a YL child faces the shock-illness of a father or mother?	-0.018	0.006	0.006	0.003	0.042***	0.042***
	(0.016)	(0.010)	(0.010)	(0.018)	(0.014)	(0.014)
Gender of a YL child	-0.102***			-0.145***		
	(0.021)			(0.023)		
Living in urban areas	0.168***			0.059*		
	(0.030)			(0.036)		
How many people, including children, live in the commune? (log value)	0.104***	0.030**	0.031**	0.019	-0.107***	-0.104***
	(0.012)	(0.014)	(0.014)	(0.014)	(0.019)	(0.019)
Problem: Is local families' garbage dumped at these water sources?	-0.036**	0.005	0.005	-0.061***	-0.031**	-0.030**
	(0.016)	(0.011)	(0.011)	(0.019)	(0.015)	(0.015)
Problem: uses of pesticides and fertilizers in local agricultural lands and faces the problem of standing water, open drains?	-0.048**	0.001	0.002	-0.144***	-0.062***	-0.060***
	(0.019)	(0.013)	(0.013)	(0.022)	(0.017)	(0.017)
Service availability in a community: Children's playground	-0.049***	0.019*	0.020*	0.027	0.037***	0.040***
	(0.015)	(0.010)	(0.010)	(0.018)	(0.013)	(0.013)
Whether the locality have any factories that employ their resident?	0.075***	0.011	0.011	0.036*	0.099***	0.099***
	(0.018)	(0.012)	(0.012)	(0.020)	(0.016)	(0.016)
Constant	-2.887***	-3.902***		0.533**	2.450***	
	(0.214)	(0.201)		(0.238)	(0.280)	
R^2	0.10	0.04	0.04	0.11	0.12	0.11
Number of observations	21,387	21,387	21,303	21,336	21,336	21,247

Notes: * $p<0.1$; ** $p<0.05$; *** $p<0.01$. Clustered standard errors at the individual level are shown in parentheses. Time and country are fixed.

As discussed in the "*Estimated results*", the estimates from the Pooled OLS method could be biased due to the influence of time-invariant unobserved variables. The FE and FEIV methods could overcome the endogeneity problem of the treatment variable. However, the Hansen test from the FEIV estimator suggests that the FE specification is appropriate to estimate the causal effect of piped water accessibility on child health. This section will interpret the estimated impacts of some important control variables on the health outcomes using the FE estimator.

The detailed results of the FE method are presented in columns 2 and 5 of Table S1. Accordingly, it is anticipated that the height of YL children will fall if they live in families with a greater child-to-household size ratio. The large families, particularly those living in impoverished areas, may experience greater financial stress, which can be detrimental to the children's health. The BMI model does not provide a significant estimate of the child-to-household size ratio. Regarding the age of the mother, we find a U-shaped relationship in the BMI model. The height model suggested that younger mothers have taller children than others.

Food and non-food expenditures have beneficial and statistically significant effects on child health, according to both models. If a household spends an additional $1 per capita on food, the child's height-for-age z-score and BMI-for-age z-score will increase by 0.044 and 0.082 units, respectively. Our results align with those of Humphries *et al.* (2017). The food spending is derived from the total amount moms spent on 21 to 33 food categories over the previous two weeks, including gifts and food products from their own businesses (Humphries *et al.*, 2017; Marion, 2018). However, we are unable to answer the question of how spending on different food groups affects child health. Using the YLS data, Weingarten *et al.* (2020) discovered that expenditures on fats had stronger relationships with the subsequent heights of Ethiopian and Indian children. Moreover, their study revealed that starch consumption had a negative effect on the future height of Vietnamese children.

We also find significant and positive effects of non-food expenditures on child health. According to Marion (2008), non-food household consumption comprises expenditures on education, health, clothes, and footwear, as well as other non-food items (e.g. rents, electricity, gas, and so on). Interestingly, the magnitude of the impact of non-food expenditures is less than that of household food consumption. The greater influence of food consumption requires governments to increase financial incentives to assure the availability of nutritious food and stable pricing in low-income regions. Children are at the greatest risk for all forms of malnutrition when they live in low-income communities, as a result of residents' preference for low-cost, low-quality food (UNICEF, 2019).

Regarding non-food household expenditures, the positive impact on children's health proposes policy solutions for supporting regional economic development, particularly in disadvantaged regions. As a result, children will have more access to medical care when they are ill and will gain indirect health advantages from improved educational access. As a substantial component of non-food expenditures-health expenditures, developing countries should implement universal health care to ensure that all citizens have access to the health services they require without incurring financial hardship. When the influence of consumer's durable index, a measure of the household's ownership of common household products such as mobile phone and television, on child health is positive and statistically significant, regional economic development policy recommendations are also reasonable. In developing nations, a rise in the ownership rates of items such as cell phones, televisions, and radios will dramatically improve mothers' awareness of safe drinking water and sanitation.

Regarding the community's characteristics, children from more polluted communities have a lower BMI. Particularly, if a child live in a community with the problem of dumping local families' garbage at local water sources, he/she will have a BMI 0.031 units lower than other children. The figure for the problem of using pesticides and fertilizers in local agricultural lands and facing the problem of standing water and open drains is -0.062 in the BMI model, statistically significant tested at any conventional levels. However, the height model provides insignificant estimates of these variables.

We also find the positive impact of service availability in a community on the health of children. Particularly, if children's playground is available in the community, children will be expected to have a higher height and BMI indicators. The estimates are statistically significant in both models. In addition, the localities having any factories that employ their residents will have a positive and statistically significant impact on child BMI. The finding also supports the health benefits of regional economic development policies in developing countries.

References

Humphries, D. L., Dearden, K. A., Crookston, B. T., Woldehanna, T., Penny, M. E., & Behrman, J. R. (2017). Household food group expenditure patterns are associated with child anthropometry at ages 5, 8 and 12 years in Ethiopia, India, Peru and Vietnam. *Economics and Human Biology*, *26*, 30–41. https://doi.org/10.1016/j.ehb.2017.02.001

Marion, P., & Lives, Y. (2018). *A Guide to Young Lives Rounds 2 to 5 Consumption Aggregates*. www.younglives.org.uk

UNICEF. (2019). *Children, food and nutrition: growing well in a changing world.*

Weingarten, S. E., Dearden, K. A., Crookston, B. T., Penny, M. E., Behrman, J. R., & Humphries, D. L. (2020). Are household expenditures on food groups associated with children's future heights in Ethiopia, India, Peru, and Vietnam? *International Journal of Environmental Research and Public Health*, *17*(13), 1–23. https://doi.org/10.3390/ijerph17134739

Table S3: The first-stage regression results of the FEIV in Table 4

	Height-for-age z-score	BMI-for-age z-score
Dependent variable:	Piped water accessibility	
The leave-out ratio in piped water source	0.741*** (0.028)	0.738*** (0.028)
Age of children in months (Log form)	-0.002 (0.035)	0.006 (0.035)
Child to household size ratio	-0.007 (0.028)	-0.002 (0.028)
The percentage of women in each household	-0.065* (0.034)	-0.074** (0.033)
Age of mother	0.004 (0.006)	0.003 (0.006)
Age of mother square	0.000 (0.000)	0.000 (0.000)
Food expenditure per capita in real 2006 price, converted to USD (log value)	0.008 (0.005)	0.008 (0.005)
Non-food expenditure per capita in real 2006 price, converted to USD (log value)	0.012** (0.005)	0.011** (0.005)
Consumer durable index	0.030 (0.024)	0.035 (0.024)
Have a household faced any natural disasters since the previous round?	0.006 (0.005)	0.006 (0.005)
Whether a YL child faces the shock-illness of a father or mother?	0.005 (0.005)	0.005 (0.005)
How many people, including children, live in the commune? (log value)	-0.005 (0.007)	-0.006 (0.007)
Problem: Is local families' garbage dumped at these water sources?	-0.002 (0.006)	-0.003 (0.006)
Problem: uses of pesticides and fertilizers in local agricultural lands and faces the problem of standing water, open drains?	-0.004 (0.006)	-0.003 (0.007)
Service availability in a community: Children's playground	-0.007 (0.005)	-0.007 (0.005)
Whether the locality have any factories that employ their resident?	-0.002 (0.005)	-0.000 (0.005)
Number of observations	21,303	21,247
Number of clusters (unique children)	6,662	6,657

Notes: * *p*<0.1; ** *p*<0.05; *** *p*<0.01. Clustered standard errors at the individual level are shown in parentheses. Time and country are fixed.

Table S4: Descriptive statistics regarding water-relevant and water-irrelevant disease symptoms

	Water-relevant disease symptoms				Water-irrelevant disease symptoms			
	the whole sample	low-educated sample (0)	educated sample (1)	Diff. (0)-(1)	the whole sample	low-educated sample (0)	educated sample (1)	Diff. (0)-(1)
Ethiopia	0.065	0.075	0.06	0.014	0.02	0.017	0.025	-0.008
India	0.141	0.150	0.134	0.016	0.016	0.015	0.017	-0.002
Vietnam	0.008	0.008	0.009	-0.000	0.007	0.013	0.007	0.006

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table S5: Attrition bias

	Sample: WITHOUT children moved out to other communities					
	Height	BMI	Height (educated sample)	Height (less-educated sample)	BMI (educated sample)	BMI (less-educated sample)
Piped water accessibility	0.014	0.019	0.034*	-0.048	0.008	0.004
	(0.016)	(0.024)	(0.019)	(0.030)	(0.028)	(0.045)
R^2	0.04	0.12	0.04	0.05	0.08	0.21
N	21,387	21,336	14,602	6,710	14,567	6,698
	Sample: WITH children moved out to other communities					
Piped water accessibility	0.007	0.021	0.024	-0.048*	0.015	-0.013
	(0.015)	(0.022)	(0.018)	(0.029)	(0.026)	(0.042)
R2	0.04	0.11	0.04	0.05	0.08	0.21
N	24,131	24,030	16,531	7,506	16,438	7,502

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table S6: On-premises piped water accessibility and height-for-age z-score: by age

	Height-for-age z-score				Height-for-age z-score			
	5 years old	8 years old	12 years old	15 years old	5 years old	8 years old	12 years old	15 years old
Piped water accessibility	0.599***	0.393***	0.563***	0.191***	0.144**	0.038	0.079**	-0.000
	(0.147)	(0.115)	(0.095)	(0.067)	(0.056)	(0.048)	(0.038)	(0.031)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.12	0.14	0.10	0.10	0.14	0.15	0.12	0.11
Observations	4,212	4,169	6,671	6,334	4,212	4,169	6,672	6,334
The first stage: Dependent variable is piped water accessibility								
The leave-out ratio in piped water source	0.802***	0.847***	0.806***	0.855***				
	(0.047)	(0.038)	(0.030)	(0.023)				
Cragg-Donald Wald F statistic	703.624	887.182	1327.225	1681.707				
Kleibergen-Paap rk Wald F statistic	285.490	490.304	726.235	1374.725				
p value of F test of excluded IVs	0.000	0.000	0.000	0.000				
Hausman test (p value)	0.0006	0.0009	0.000	0.0013				

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage is pipedw, whether household access to piped water into dwelling/yard/plot? Instruments include both selected leave-out commune ratios. Regarding the STATA command, we used ivreg2. Control variables reflect the characteristic of children, their families, and their communities. We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported because of the exactly identified regression model.

Table S7: On-premises piped water accessibility and BMI-for-age z-score: by age

	BMI-for-age z-score				BMI-for-age z-score			
	5 years old	8 years old	12 years old	15 years old	5 years old	8 years old	12 years old	15 years old
Piped water accessibility	0.869***	0.564**	0.294**	0.053	0.138**	0.016	0.026	-0.067
	(0.182)	(0.240)	(0.118)	(0.092)	(0.062)	(0.056)	(0.046)	(0.041)
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family’s characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.10	0.08	0.15	0.16	0.13	0.10	0.16	0.17
Observations	4,192	4,146	6,671	6,326	4,192	4,146	6,672	6,326
The first stage: Dependent variable is piped water accessibility								
The leave-out ratio in piped water source	0.790***	0.846***	0.805**	0.857***				
	(0.048)	(0.028)	(0.029)	(0.027)				
Cragg-Donald Wald F statistic	682.531	877.896	1441.107	1765.779				
Kleibergen-Paap rk Wald F statistic	275.230	936.916	775.301	1426.076				
p value of F test of excluded IVs	0.000	0.000	0.000	0.000				
Hausman test (p value)	0.000	0.018	0.018	0.220				

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage is pipedw, whether household access to piped water into dwelling/yard/plot? Instruments include both selected leave-out commune ratios. Regarding the STATA command, we used ivreg2. We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported because of the exactly identified regression model.

Table S8: On-premises piped water accessibility and child anthropometric outcomes: by gender

	Height-for-age z-score				BMI-for-age z-score			
	Boys		Girls		Boys		Girls	
	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE
Piped water accessibility	0.117 (0.072)	0.025 (0.021)	-0.026 (0.081)	0.001 (0.025)	0.169 (0.125)	-0.006 (0.034)	0.136 (0.108)	0.022 (0.031)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Covariates								
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family's characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Community characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.04	0.05	0.06	0.06	0.16	0.16	0.12	0.12
Observations	11,215	11,252	10,088	10,135	11,094	11,135	10,153	10,201
The first stage: Dependent variable is piped water accessibility								
The leave-out ratio in piped water source	0.737*** (0.037)		0.746*** (0.043)		0.728*** (0.037)		0.750*** (0.043)	
Cragg-Donald Wald F statistic	766.438		673.071		740.668		680.487	
Kleibergen-Paap rk Wald F statistic	388.202		304.158		380.324		310.656	
p value of F test of excluded IVs	0.000		0.000		0.000		0.000	
Hausman test (p value)	0.175		0.727		0.141		0.275	

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Notes: Standard errors shown in parentheses are clustered by a child level. The dependent variable in the first stage of the FEIV approach is pipedw, whether household access to piped water into the dwelling/yard/plot? We solely run the FEIV method with an IV which is the leave-out commune ratio in piped water sources on the premises. Sargan-Hansen test is not reported because of the exactly identified regression model.